Proceedings of the

1998 Conference on the

History and Heritage of

Science Information Systems

Generously Supported By:

The Chemical Heritage Foundation

The Eugene Garfield Foundation

The National Science Foundation (Grant #IIS-9814034)

Proceedings of the

1998 Conference on the

History and Heritage of

Science Information Systems

Edited by

Mary Ellen Bowden

Trudi Bellardo Hahn

Robert V. Williams

ASIS Monograph Series

Published for the
American Society for Information Science and the Chemical Heritage Foundation by

Information Today, Inc.

Medford, NJ
1999

Printed in the United States of America.

Published by
Information Today, Inc.
143 Old Marlton Pike
Medford, NJ 08055-8750

Distributed in Europe by
Learned Information Europe, Ltd.
Woodside, Hinksey Hill
Oxford OX1 5BE
England

To contact CHF write
Chemical Heritage Foundation
315 Chestnut Street
Philadelphia, PA 19106-2702, USA
Fax: (215) 925-1954

Library of Congress Cataloging-in-Publication Data

Conference on the History and Heritage of Science Information Systems (1998 : Pittsburgh)
Proceedings of the 1998 Conference on the History and Heritage of Science Information Systems / edited by Mary Ellen Bowden, Trudi Bellardo Hahn and Robert V. Williams.
p. cm. — (ASIS monograph series)
Includes bibliographical references and index.
ISBN 1-57387-080-3 (pbk.)
1. Information storage and retrieval systems—Science—History Congresses. 2. Information storage and retrieval systems—Bibliography—History Congresses. I. Bowden, Mary Ellen. II. Bellardo, Trudi. III. Williams, Robert Virgil, 1938– . IV. Title. V. Series.
Z699.5.S3C64 1998
025.06′5—dc21 99-36330
CIP

Cover design: Jacqueline Walter

To

the pioneers

of

science

information

Contents

Foreword

The Chemical Heritage Foundation and the American Society for Information Science convened the first conference on the history and heritage of science information systems in October 1998. The conference, which examined the historic roles of the chemical sciences and of chemists in the development of information systems, among many other topics, greatly benefited from the energy and generosity of one chemist and information scientist, Eugene Garfield.

Gene Garfield stands in a long line of chemists and information innovators. That line begins with Robert Boyle and the organization of the Royal Society and its *Philosophical Transactions* in the seventeenth century and Antoine Lavoisier and the reform of chemical nomenclature and the creation of the *Annales de chimie* late in the eighteenth century. In more recent times chemical giants like Wilhelm Ostwald and J. D. Bernal have been great visionaries of science information systems, while other chemists like James W. Perry and Frederick A. Tate have taken the lead in designing and using increasingly sophisticated automated systems. In Gene Garfield's case the very products of his enterprise are of immediate use to the historian of science as well as to the scientist. *Current Contents*, the *Science Citation Index*, and other similar tools from the Institute for Scientific Information (ISI) allow the historian to identify the members of schools of thought and trace the growth of these schools and the growth of whole fields of science.

It was through a mutual interest in these quantitative measures of science, scientometrics, that I first met Gene when we both became involved in 1970 in the effort to launch the Society for the Social Studies of Science (4S). The list of winners of the 4S's Bernal Award that he initiated is a group of sociologists and historians worthy of Nobel Prizes, including such luminaries as Derek Price and Robert K. Merton. Gene has also proved to be an enthusiast for a more traditional kind of history: the biographical memoir. In addition to four thousand *Citation Classic* bibliographical commentaries, he often chose to use the editorials in *Current Contents* to honor great scientists, including pioneer information scientists—about whom little or nothing had been written. The references cited in Gene's paper in this volume can only hint at the extensiveness of his own historical writings, which are posted on his home page at http://garfield.library.upenn.edu/index.html.

To involve others in his love for information science and its history, Gene has funded a growing web of activities at the Chemical Heritage Foundation. CHF's Eugene Garfield Fellowship in the History of Science Information has stimulated numerous oral histories of information science pioneers and a chronology of chemical information science. Gene also generously supported the CHF/ASIS conference, which in turn gave rise to this volume to serve as inspiration for future historical and policy-oriented research.

Arnold Thackray, President
Chemical Heritage Foundation
30 July 1999

Preface

The Conference on the History and Heritage of Science Information Systems, held 23–25 October 1998, in Pittsburgh, Pennsylvania, brought to fruition the efforts of a wide variety of people. Over the last few years a small band of enthusiasts has determinedly pursued the history and heritage of science information, even though there was little support and only rare appreciation of this historical enterprise. Scholars working abroad or those outside the field of information science—such as the few historians of science and technology who had approached this field in their investigations—received even less support (monetary or otherwise). These individuals usually found themselves isolated geographically from like-minded individuals or separated by disciplinary boundaries. The original purpose of the conference was to bring together as many of these dedicated people as possible to share with each other their research, insights, and knowledge. The conference organizers also seized the opportunity, unusual in most historical exercises, of inviting the historical figures themselves—pioneers in the automation of science information—to contribute papers or simply bear witness to the past in the form of brief reminiscences. This volume shows, I believe, just how marvelously eclectic was the conference, how stimulating was the exchange of views, and what exciting opportunities for future research exist.

Among information scientists, the origins of this conference go back to the decision by the American Society of Information Science (ASIS) Foundations of Information Science Special Interest Group to reformulate the group as the *History* and Foundations of Information Science and to organize special sessions at annual meetings. The new group requested funds from ASIS to identify and document the contributions made by the pioneers of information science in North America over the last hundred years. Supported by ASIS, I was able to compile the desired information and build a Web database (www.asis.org/Features/Pioneers/isp.htm). It became a much bigger job than I ever anticipated, but at the same time, it also became a labor of love. Meanwhile, Boyd Rayward, Michael Buckland, and Trudi Bellardo Hahn launched a series of editorial projects to construct historical bibliographies and bring together history papers in the field—most recently Buckland and Hahn's *Historical Studies in Information Science* (1998).

I came into contact with historians of science and technology, particularly those of the chemical persuasion, and their desires for a conference as a result of my calling Eugene Garfield, one of the preeminent pioneers of information science. I called him to find out more about the disposition of his papers and the archives of the Institute for Scientific Information. He told me that he had recently done an oral history interview with the Chemical Heritage Foundation (CHF). A call to Mary Ellen Bowden, senior research historian at CHF, led not only to information about Garfield's oral history interview but also to the discovery of Gene's imaginative and generous decision to launch, through CHF, the Garfield Fellowship in the History of Science Information. A few months later I was CHF's first Garfield Fellow (1997–98).

Robert V. Williams, Coeditor

Acknowledgments

In his stimulating "an idea every day" approach to leadership, Arnold Thackray, president of CHF, had formulated ideas for a conference on the history of science information long before I arrived at CHF headquarters. He soon had me developing a plan and forming a planning committee. I was fortunate to entice key researchers and proponents of the history of information science to serve on the committee that planned the conference and peer-reviewed submitted papers: Michael Buckland, Colin ("Brad") Burke, Toni Carbo, Irene Farkas-Conn, Eugene Garfield, Trudi Bellardo Hahn, and Boyd Rayward. My thanks to all their efforts. I would also like to mention the CHF staff members who lent their expertise to the committee—Mary Ellen Bowden, Leo Slater, Marie Stewart, and Arnold Thackray—and Dick Hill, executive director of ASIS, who, as an ex-officio member, smoothed the way in all areas related to hotel and other meeting arrangements. The committee worked diligently for a full sixteen months before the conference, meeting irregularly in Philadelphia and regularly via conference calls.

My work as the Garfield Fellow, which involved conducting oral histories as well as preparing chronologies of chemical information and information science and technology, drew upon the energies of the entire CHF staff. When I returned to my "real job" of full-time teaching, the challenge of supporting these endeavors over time and distance became immeasurably more difficult. I particularly wish to acknowledge the CHF staff members most directly involved in planning, scheduling, and arranging the conference: June Bretz, Laura Myers, Janine Pollock, Marie Stewart, and Monica Womack.

Converting conference papers into published papers required the knowledge and wisdom of my two coeditors, Trudi Hahn and Mary Ellen Bowden, who served as the linchpin for this effort; a corps of copyeditors, led by CHF's Shelley Wilks Geehr and Patricia Wieland; and the cheerful cooperation of our authors in making requested revisions.

Finally, historical enterprises usually suffer from a lack of funding. But in our case we received additional generous support from the Eugene Garfield Foundation, as well as from CHF and the National Science Foundation. Without the support of these organizations the conference might never have happened nor might these proceedings reach the wider audience in whom we wish to kindle a pride in the history and heritage of science information, a desire to preserve essential documents and artifacts, and—at least among a few readers—the desire to analyze and write about that past.

Robert V. Williams, Coeditor
Chair, Conference Planning Committee
University of South Carolina, College of Library and Information Science

Contributors

Geoffrey C. Bowker
Associate Professor
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

Mark D. Bowles
Graduate Student
Case Western Reserve University
Cleveland, Ohio

Michael Buckland
Professor
School of Information Management and Systems
University of California, Berkeley

Susan A. Cady
Group Leader
Administration, Planning and Advancement Services
Lehigh University Information Resources
Bethlehem, Pennsylvania

Bernadette G. Callery
Museum Librarian
Carnegie Museum of Natural History
Pittsburgh, Pennsylvania

James M. Cretsos
Retired
Cincinatti, Ohio

Bernd Frohmann
Associate Professor
Faculty of Information and Media Studies
University of Western Ontario
London, Ontario

Eugene Garfield
Chairman Emeritus
Institute for Scientific Information
Publisher
The Scientist
Philadelphia, Pennsylvania

Ruggero Giliarevskii
Professor and Head
Informatics Theory Department
All-Russian Institute for Scientific and Technical Information (VINITI)
Moscow, Russia

Thomas Hapke
Subject Librarian for Chemical Engineering
Technical University
Hamburg-Harburg, Germany

Robert M. Hayes
Professor Emeritus
University of California, Los Angeles

Madeline M. Henderson
Retired
Mechanicsville, Maryland

Thomas P. Hughes
Professor Emeritus
Department of History and Sociology of Science
University of Pennsylvania, Philadelphia

Florence H. Kvalnes
Librarian
Corporate Information Science
Central Research and Development
E.I. DuPont de Nemours
Wilmington, Delaware

Timothy Lenoir
Professor
Program in History and Philosophy of Science
Stanford University
Palo Alto, California

Bruce V. Lewenstein
Associate Professor
Departments of Communication and Science and Technology Studies
Cornell University
Ithaca, New York

Herbert Ohlman
Retired
Mount Pleasant, South Carolina

Jill O'Neill
Writer
Institute for Scientific Information
Philadelphia, Pennsylvania

Pamela Spence Richards
Professor
School of Communication, Information, and Library Studies
Rutgers University
New Brunswick, New Jersey

Takashi Satoh
Professor
University of Library and Information Science
Tsukuba, Japan

Helen Schofield
Chemistry Librarian and Honorary Lecturer
University of Manchester
Institute of Science and Technology
Manchester, United Kingdom

Robert W. Seidel
Professor, History of Science and Technology Program
Department of Chemical Engineering and Materials Science
University of Minnesota, Minneapolis

Jacqueline Trolley
Director, Corporate Communications
Institute for Scientific Information
Philadelphia, Pennsylvania

Jana Varlejs
Professor
School of Communication, Information, and Library Studies
Rutgers University
New Brunswick, New Jersey

Paul Wouters
Researcher, Nederlands Institut voor Wetenschappelijke Informatiediensten
Amsterdam, Netherlands

History

and

Historiography

of

Science Information

Systems

Overview of the History of Science Information Systems

Michael Buckland

Abstract

This overview is an introduction to the history and heritage of science information systems and a discussion of historiography of this area. History is narrative of events in time past. The continuing consequences of those events are heritage, which includes our collective memory, our understanding of history. The heritage of information systems is of additional significance because the systems' design and characteristics have long-lasting effects.

The history of science information systems overlaps the history of science, the history of information systems (previously documentation), and the history of technology. It includes the usual genres of historical inquiry: biography, archaeology, cliometrics (here, especially, bibliometrics and infometrics), oral history, and documentary research, with their differing strengths and weaknesses.

Information systems for science and technology have had a privileged existence because of industrial and military needs and government policies. Much of the pioneering work in science information systems was concerned with chemistry or pioneered by individuals trained in chemistry.

The past decade of work on the history and heritage of information systems is summarized. Several initiatives have been undertaken to encourage research and to build a supportive infrastructure, which is important if historical research in this area is to be sustained and to flourish. This conference is itself a significant part of that effort.

Welcome

It is an honor and a pleasure to welcome you all in my capacity as president of the American Society for Information Science, founded more than sixty years ago in 1937 under the name American Documentation Institute to advance the development of information systems and services. In addition, I extend a welcome from the ASIS Special Interest Group on the History and Foundations of Information Science, which, in the past several years, has nurtured attention to the history of information science.

My remarks are mainly concerned with the history of information systems and services generally. However, science information systems have had a privileged status because of industrial and military needs and government policy and also perhaps because the domains of science appear more tractable for information systems than in the social sciences and humanities.

Much of the pioneering thought and work in the development of information systems was first done in relation to chemistry or by chemists. Among individuals one thinks of Wilhelm Ostwald and Emanuel Goldberg in Germany and of Frits Donker Duyvis in the Netherlands. Among historically important information centers one thinks of the Maison de Chimie in France and, of course, of *Chemical Abstracts* in the United States.

Information Systems as a Typical Field for Historians

The history of information systems has the usual features, genres, and specialties as other fields of historical study. There are biographies of diverse kinds. Boyd Rayward's biography of Paul Otlet is a notable example of a biography of a person (Rayward, 1975). Irene Farkas-Conn's study *From Documentation to Information Science* is a biography of an organization, the American Society for Information Science in its early years (1990). Colin Burke's reconstruction of the development of the microfilm rapid selector (and the related comparators) by Vannevar Bush is a good example of a biography of a machine (1994).

As an example of a kind of archaeology, I cite the short documentary by Robert Williams on Termatrex optical coincidence retrieval technology (Williams & Covey, 1990).

Cliometrics, quantitative historical analysis, is well represented by a recent study by Howard White and Katharine McCain (1998), who used co-citation analysis to illuminate the development of the field from 1972 to 1995. Their analysis reveals a field composed of two large groups, remarkably stable and remarkably separate from each other for twenty-four years.

The Chemical Heritage Foundation and the Eugene Garfield Foundation have recently supported oral history of information systems with pioneers of chemical information systems in the form of Robert Williams's interviews.

Intellectual and cultural history is present here, too, not least in the tensions before World War II between documentation and librarianship and similarly after World War II, between librarianship and information science. These were significant but complex phenomena still far from understood (Buckland, 1996; Fayet-Scribe, 1997; Williams, 1997).

In the history of information systems, like any other field, we have our mythic history, narratives that are even more mythic than history ordinarily is. The Memex phenomenon, with the engineer-administrator Vannevar Bush as an icon, is a good example. Bush is rightly famous. He led the technology effort for World War II, creating the atomic age, and was the father of the National Science Foundation (Zachary, 1997). Yet he is best known in the field of information retrieval, even though his systems hardly worked, his ideas were not new, he did not really understand what he was talking about, and he chose not to acknowledge the priority of others (Fairthorne, 1958; Buckland, 1992; Zachary, 1997, p. 265). Nevertheless the citing of Bush's 1945 essay "As We May Think" has been so intense that the citing itself has become an object of research (Bush, 1945; Smith, 1991). For some, such as Doug Engelbart and Ted Nelson, this well-written essay was unquestionably a genuine, powerful, and productive inspiration. Bush, however, was not just any author; he was the "engineer of the American century," the engineer-administrator who epitomized success. To associate oneself with Bush by linking one's own writings to his was to claim legitimacy and respectability among peers and funders. So, for others, invoking Bush's Memex was in effect a self-interested political gesture. J. C. R. Licklider, who was very successful in this environment, effusively dedicated his book *Libraries of the Future* to Bush, citing "As We May Think" as the "the main external influence that shaped the ideas of this book," even though he had not read Bush's essay until after the book had been written (Licklider, 1965, pp. xii–xiii). Still other writers seem to have cited Bush because everyone else seemed to be doing so.

That the invocation of Bush was driven by social and political, as much as intellectual considerations, is confirmed by the ahistorical positioning of Bush. Memex is usually cited in isolation. Associating one's work with others without Bush's aura would not have had the same attraction in the competitive positioning in U.S. science and engineering. If the purpose of citing were simply to acknowledge priority, then others such as Paul Otlet and Emanuel Goldberg, who had anticipated Bush's ideas, would have been mentioned. They were, however, dead by then or far removed from the sources of power in the academic-government-industrial complex in the United States. They were ignored and forgotten until resurrected by writers concerned with history, while Bush's work continues to be celebrated.

If Bush had little direct part in the history of the development of information systems, he has had a very large part in the heritage of the field. This conference is very properly concerned with heritage as well as history. Heritage is what is passed down, what is perceived by each generation to be its origins and its culture. History, which consists of attempts to create narratives of what actually happened, is a part of the heritage.

Heritage has special significance in technical fields because techniques and technology have lingering effects. Once an information system has been adopted, there is a vested interest in it, and little opportunity may be left for alternative designs. Information systems, once adopted, create legacies. We have to live with the consequences of the data collection, data categorization, and data-processing decisions of the past because it is impossible or unaffordable to make retroactive changes. Even the adoption of improved practices is inhibited because changes could create incompatibilities or inconsistencies with the inherited data and systems.

An Unusual Relationship: History and Information Systems

Anyone concerned with information systems must necessarily be interested sooner or later in information. And, for anyone interested in information, history has a special attraction because history is concerned with analyzing, weighing, and interpreting the available evidence, especially documentary evidence. Information systems are concerned with the selection, representation, and preservation of available evidence, especially documents. "No documents, no history," wrote the historian Fustel de Coulanges, but the creation, survival, and accessibility of documents is an accident-prone matter. So is their content. Consider, for example, oral history transcripts, sometimes the best or only available documentary source for past events. The content and shape of the reminiscences are influenced by many factors, including how the interviewer posed questions. The spoken words are more or less edited in the creation of transcripts. When recording oral narratives, one can almost see the story being constructed as the narrator strives to make sense of what is remembered of what happened long ago. Verbatim quotations from fifty years ago are liable to come out differently worded at different times. This does not invalidate what is recorded but rather requires one to respect what they are—informed reminiscences. They are themselves a form of history—partial narratives. Collecting oral history should be part of the apprenticeship of every historian. The whole process is highly accidental: who survived, what they knew, what they recalled, what they imagined, what they chose to relate, how they chose to express it, and of course whether anybody bothered to record them. Oral histories depend not only on frail memories but also on the happenstances of who survives to tell their tale and whether anyone is around motivated enough to record them. With oral history, one is conscious of how accidental is the writing of history.

Documentary resources are similarly accident prone: Wars, fires, floods, modesty, shortages of space, and many other factors cause documents to be lost. What history is written will depend on whose papers are kept, whose have been destroyed, who cares enough to read them, whether they can be found, and how well they are understood. In the writing of history it is not only a question of which sources are to be privileged by the historian but also which sources are available to be privileged or have been privileged by the information systems professionals responsible for selecting, collecting, retaining, and representing them.

Historians and Pioneers

A historian is someone who narrates an account of what happened in some past event. At this conference we are using the term *pioneer* to refer to those who were there, who participated in those past events. This usage reminds us that historians are ordinarily people who were not there when the events they describe took place. We are very pleased that several pioneers of science information systems have been able to attend this conference. Even better, some of them will be presenting papers, performing the role of historian as well as that of pioneer.

Antecedents

The emergence of a systematic body of history of information systems is a recent development, and this conference has some important, direct antecedents. Up to 1991 there had been little attention paid to the history of information systems. In 1991 a few people decided to do something about it. They organized a historical session at the annual meeting of the American Society for Information Science titled "Information Science before 1945," and a session has been associated with each annual meeting since. These sessions, organized by, among others, Irene Farkas-Conn, Trudi Bellardo Hahn, and Robert Williams, have provided a forum for discussion and have encouraged the development of a community of interest.

Creating a community is like gardening. You cannot make plants grow, but the growth of plants can be helped or hindered. The nurturing of a community interested in the history of information systems has been consciously cultivated by a series of steps taken, largely within or through the American Society for Information Science, to build a supportive infrastructure. An initiative by Robert Williams to establish a Special Interest Group for history resulted in the expansion of an existing group concerned with theory to form the present Special Interest Group on the History and Foundations of Information Science. It seemed wise that those concerned with ideas should be historically informed and historians should be encouraged to address the history of ideas.

Another investment in infrastructure was the creation of a database of pioneering individuals and organizations: who they were and what was known about the location of their personal and professional papers. The idea was that identifying both research-worthy targets and documentary resources would not only facilitate the work of those already active in the history of information science but would also encourage historians in adjacent areas to broaden their interests to include the history of information systems. Under Williams's leadership the Pioneers of Information Science in North America database came into being (Williams, 1998).

Understanding of the history of this field has been inhibited by the lack of a systematic guide to existing writings. Therefore, a survey was prepared, published in the *Annual Review of Information Science and Technology* for 1995 and recently updated (Buckland & Liu, 1995, 1998).

Special issues of existing journals provide a forum and help to build a community. Both the call for papers and the papers themselves receive wide attention. W. Boyd Rayward guest-edited a special issue of *Information Processing and Management* in 1996 with six substantial articles (Rayward, 1996). This was followed in 1997 by a two-part historical issue of the *Journal of the American Society for Information Science*, containing fourteen articles and two bibliographies (Buckland & Hahn, 1997). The authors were from eight different countries and, I surmise, had been largely unaware of each other's work. One of the pleasures of recent years has been encountering individuals with an existing interest in the history of information science who had been toiling more or less in isolation.

Meanwhile, the Chemical Heritage Foundation, with the help of the Eugene Garfield Foundation, has been supporting oral history work by Robert Williams among pioneers of chemical information systems: Dale Baker, Melvin S. Day, Eugene Garfield, Madeleine Berry Henderson, Saul Herner, and Claire Schultz (Hahn & Buckland, 1998, p. 180).

In the absence of a textbook on the history of information systems the next best thing seemed to be a volume reprinting a selection of the recent research literature, with some new material. Preparation of this volume, *Historical Studies in Information Science*, has been timed for it to become available at this conference (Hahn & Buckland, 1998). ASIS, Wiley, and the editors and authors waived royalties, and Elsevier charged less than its standard fees for reprinting.

In this way a small but growing international research community is beginning to emerge. We hope that this group will continue to grow and become a viable self-sustaining community.

This Conference

This conference is planned to be more than an opportunity for a small community to come together. The intent is to build a broader community. The conference itself is a way to hoist the flag, a way to tell people engaged in the history of science, the study of science practice, the history of technology, the history of computing, and other neighbors that we are here. The message, however, is not only to assert the existence of this field but to reach out. We have invited speakers from outside. We are inviting neighbors in, in order to build a broader community.

The year 1998 is auspicious in that it is the anniversary of two major milestones. It is the fiftieth anniversary of the Royal Society Scientific Information Conference held in London in 1948 (Royal Society, 1948). It is also the fortieth anniversary of the International Conference on Scientific Information, sponsored by the National Science Foundation, the National Academy of Sciences, and the American Society for Information Science, previously named the American Documentation Institute (National Academy of Sciences, 1959). This conference is cosponsored by the Chemical Heritage Foundation and by the American Society for Information Science, and we are very grateful to the Eugene Garfield Foundation and the National Science Foundation for their encouragement and financial support. The National Science Foundation grant includes an obligation to plan what steps to take next.

Let us hope that in the future, forty or fifty years hence, people will look back on the 1998 conference in Pittsburgh as a milestone as significant in its way as those of 1948 and 1958.

References

Buckland, M. K. (1992). Emanuel Goldberg, electronic document retrieval, and Vannevar Bush's Memex. *Journal of the American Society for Information Science, 43*, 284–294.

Buckland, M. K. (1996). Documentation, information science, and library science in the U.S.A. *Information Processing & Management, 32*, 63–76. Reprinted in Hahn & Buckland (1998), 181–192.

Buckland, M. K., & Hahn, T. B. (Eds.) (1997, April & September). Special topic issue: History of documentation and information science. *Journal of the American Society for Information Science, 46*, 285–288, 775–842.

Buckland, M. K., & Liu, Z. (1995). History of information science. *Annual Review of Information Science and Technology, 30*, 385–416.

Buckland, M. K., & Liu, Z. (1998). History of information science. In T. B. Hahn & M. Buckland (Eds.), *Historical studies in information science* (pp. 272–295). Medford, NJ: Information Today. (This is an updated version of Buckland & Liu, 1995).

Burke, C. (1994). *Information and secrecy: Vannevar Bush, ultra, and the other Memex*. Metuchen, NJ: Scarecrow Press.

Bush, V. (1945, July). As we may think. *Atlantic Monthly, 176*, 101–108.

Fairthorne, R. A. (1958). Automatic retrieval of recorded information. *Computer Journal, 1*, 36–41. Reprinted in R. A. Fairthorne (1961). *Towards information retrieval* (pp. 135–146). London: Butterworths.

Farkas-Conn, I. S. (1990). *From documentation to information science: The beginnings and early development of the American Documentation Institute–American Society for Information Science*. New York: Greenwood.

Fayet-Scribe, S. (1997, September). The cross-fertilization of the U.S. public library model and the French documentation model through the French professional associations between World War I and World War II.

Journal of the American Society for Information Science, 46, 782–793. Reprinted in Hahn & Buckland (1998), 181–192.

Hahn, T. B., & Buckland, M. K. (Eds.) (1998). *Historical studies in information science*. Medford, NJ: Information Today.

Licklider, J. C. R. (1965). *Libraries of the future*. Cambridge, MA: MIT Press.

National Academy of Sciences–National Research Council (1959). *Proceedings of the International Conference on Scientific Information: Washington, DC—November 16–21, 1958*. Washington, DC: National Academy of Sciences, National Research Council.

Rayward, W. B. (1975). *The universe of information: The work of Paul Otlet for documentation and international organisation* (FID 520). Moscow: VINITI.

Rayward, W. B. (Ed.) (1996). *Information Processing & Management, 32*, 1–88. Special issue on the history of information science.

Royal Society. (1948). *Scientific information conference (21 June–2 July 1948): Report and papers submitted*. London: Royal Society.

Smith, L. C. (1991). Memex as an image of potentiality revisited. In J. M. Nyce and P. Kahn (Eds.), *From Memex to hypertext: Vannevar Bush and the mind's machine* (pp. 261–286). San Diego, CA: Academic Press.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science, 49*, 327–355.

Williams, R. V. (1997, September). The documentation and special libraries movement in the United States, 1910–1960. *Journal of the American Society for Information Science, 46*, 782–793. Reprinted in Hahn & Buckland (1998), 173–179.

Williams, R. V. (1998). The pioneers of information science in North America. In T. B. Hahn & M. Buckland (Eds.), *Historical studies in information science* (pp. 269–271). Medford, NJ: Information Today. The database is at http://www.libsci.sc.edu/bob/isp/isp.htm.

Williams, R. V., & Covey, H. C. (1990). *The Jonkers Termatrex machine: Background and demonstration*. Columbia, SC: University of South Carolina Distance Education Division. VHS videocassette.

Zachary, G. P. (1997). *Endless frontier: Vannevar Bush, engineer of the American century*. New York: Free Press.

Funding a Revolution

Thomas P. Hughes

The following is an introduction to a National Research Council report titled *Funding a Revolution: Government Support for Computing Research* (National Research Council, 1999). The introduction represents the views of the chairperson, Thomas P. Hughes, not those of the Committee on Innovation in Computing and Communications or of the National Research Council. The final report can be ordered from National Academy Press, 2101 Constitution Avenue, NW, Box 285, Washington, DC 20055. The text of the report is now available online, and it may be ordered electronically. The URL is http://www.nap.edu/readingroom/books/far.

Introduction

At a time when the American style of competitive market capitalism attracts the world's attention, even its envy, it is difficult to recall and acknowledge that since World War II, the federal government has played a major role in launching and giving momentum to the information revolution that now takes pride of place among the nation's recent technological achievements. Federal funding financed development of most of the nation's early digital computers and, even as the industry matures, continues to finance breakthroughs in areas as wide ranging as computer time-sharing, networking, artificial intelligence, and virtual reality. The government also continues to support the education of undergraduate and graduate students who now populate industry and academic research centers and to fund the development of the physical infrastructure needed for leading-edge research.

This information revolution that the government has helped fund is not simply a technical change; it is a sociotechnical revolution, comparable to an industrial revolution. The British Industrial Revolution, for instance, which in the late eighteenth century ushered in the modern era, brought not only steam power and factories but also the rise of industrial cities and a politically powerful urban middle class and a worker class soon empowered by trade unions. The profession of civil engineering grew rapidly, applying the laws of nature to the transformation of the environment.

The sociotechnical aspects of the information revolution are now becoming clear as firms producing microprocessors and software are challenging the economic power of firms with factories manufacturing automobiles and refineries producing oil. Detroit is no longer the symbolic center of an American industrial empire; Silicon Valley, California, now conjures up visions of enormous entrepreneurial vigor. Men in board rooms and gray flannel suits are giving way to the easy manners and casual dress of young founders of start-up computer companies.

Today the information revolution continues with private companies increasingly funding research and development for computing and related communications. Yet the federal government continues to play a major role, especially by funding long-term, high-risk research. Given the successful history of federal involvement, several questions arise: Are there lessons to be drawn that can inform future policy making in the realm of research and development? What roles might the government play in sustaining the information revolution and helping to initiate other comparable technological developments? The fact that the government funding produced—and will continue to produce—social as well as technical change adds to the responsibilities of those making science and technology policy.

Funding a Revolution reviews the history of innovation in computing and communications and seeks to identify factors that have contributed to the nation's success in these fields. It presents and draws lessons from a series of case studies that trace the lineage of innovations, in particular subdisciplines of computing and communications (see box, pp. 10–11). *Funding a Revolution* also presents and seeks to draw lessons from a more general historical review of these industries since World War II. The lessons are intended to provide general guidance for those shaping current and future federal policy.

From these lessons emerge three central themes: 1) the importance of collaboration and coordination among members of the government-industry-university complex in cultivating research and development; 2) the positive results from diversity and change in the mix of federal organizations funding research and development and in their styles of research and development support; and 3) the importance of sound program management in federal agencies (see table).

Government-Industry-University Interaction

Innovation in computing and communications stems from a complementary interaction among government, industry, and universities. In this complex relationship, government agencies and private companies fund research that is conducted in a mixture of university, industry, and government laboratories. Industrial research laboratories often partner with government-funded academic research centers to conduct research and development and to generate innovations. Joint ventures, consortia, and partnerships involving government, industry, and universities have also stimulated and sustained the ongoing information revolution. These arrangements transcend the activities of individual firms that, earlier in this century, usually drew on in-house research and development.

The federal government has generally played a critical role in funding fundamental, long-term research, whereas industry tends to support research and development with more immediate and discernible market potential. At other times, however, government support has been closely tied to particular missions, whether national defense, space exploration, or health. A case in point would be government funding of military computer-based command and control systems.

Universities provide a culture conducive to fundamental research. Between 1972 and 1995, the federal government supported roughly 70 percent of university research in computer science and about 65 percent of university research in electrical engineering. Fundamental research has often found application. An example is the Project on Mathematics and Computation (MAC) at the Massachusetts Institute of Technology (MIT). Sponsored by the Defense Advanced Research Projects Agency (DARPA), this project advanced computer time-sharing techniques, demonstrated the capabilities of computer-based utilities based on time-sharing, and helped clarify many now ubiquitous notions of computer systems. Work on Project MAC also prompted the development of a simpler derivative architecture that became UNIX.

Table: Summary of Lessons*

Lessons about Government, Industry, and University Collaboration
1. Government funding of long-term, high-risk research complements the application-oriented research and development activities of industry.
2. Government (especially the military) has funded large system-building projects. In alliance with industry and universities, it has designed, researched, and developed these projects.
3. Government is the primary supporter of university research.
4. The free flow of people and ideas within the government-industry-university complex is critical to disseminating information about and spreading new styles of research and development.
Lessons about Diversity and Change in Federal Funding
5. Research and development in computing and communications has benefited from a diversity of approaches pursued by federal funding agencies and from organizational innovation among federal agencies.
6. Federal funding has supported both fundamental research and mission-oriented research and development.
Lessons about Program Management
7. Successful research and development programs require both talented researchers and nurturing environments. Gifted program managers have helped create these environments.
8. Program managers have often stimulated fruitful collaboration between university and industrial researchers.
9. Successful federal program managers have often shown a light management touch.
10. Experienced program managers have pursued policies based upon their realization that research and development is a more complicated process than the linear applied-science model suggests.
11. Federal program managers have often funded research that is inherently unpredictable. The unanticipated results have often been fruitful.

* A revised version of the "lessons" is in *Funding a Revolution,* pp. 5–13.

Diversity and Change

Diversity and change in government funding policies are characteristic of the ongoing revolution. Multiple agencies frequently provide funding for projects in related areas, often backing different technological approaches. Such diversity and change do not result from indecision or lack of focus but are a measured response to changes in the conditions that constitute the context for funding and to changes in the technology and organizations being supported.

Different funding agencies also focus on different phases of the research, development, and deployment process. Those responsible for funding policy realize that the research, development, and deployment process does not flow simply and directly from basic research, through commercial-supported applied research and development, to deployment. Federal program and project managers have to adjust, for instance, to the messy reality that mission-oriented technological development may stimulate fundamental research. Fundamental research, they learned, often rationalizes or explains technology developed earlier through cut-and-try experimentation. For example, the engineers who developed the interface message processors (IMPs), or gateway computers, for the ARPANET often found themselves advancing empirically beyond theory.

Defense agencies (notably DARPA) and the National Science Foundation (NSF) have been the primary federal supporters of research in computer science and electrical engineering, the two academic disciplines most closely related to computing and communications. Other agencies, such as the Department of Energy (DOE) and NASA, have supported work relevant to their missions. Each agency has its own style of operating. In the 1960s DARPA concentrated large research grants in what it called "centers of excellence." In time, these centers matured into some of the country's leading academic computer departments. Other federal agencies have supported individual researchers at a more diverse set of institutions. The Office of Naval Research and the NSF awarded numerous peer-review grants to individual researchers, especially in universities. The NSF has also been especially active in awarding fellowships to graduate students.

In summary, federal support takes many forms: support of basic and fundamental research, support of mission-oriented development projects, research grants to institutions and centers of excellence, research grants to individuals, fellowships for graduate students, and procurement of hardware and software.

Case Studies in Computing and Communications

The case studies are contained in chapters 4 though 9 of the published report. These histories of artificial intelligence, relational databases, computer networking, virtual reality, theoretical computer science, and very-large-scale integrated circuits demonstrate the interaction of government, universities, and industry in developing and commercializing new information technology. Though representing a range of technologies and timeframes, the cases display a number of interesting similarities and contrasts that highlight key elements of the innovation process. A summary of the histories follows.

Artificial Intelligence

Support for research in artificial intelligence (AI) over the past three decades has come largely from government agencies, such as the Defense Advanced Research Projects Agency (DARPA), the National Science Foundation (NSF), and the Office of Naval Research (ONR). Firms that initiated AI research programs in the 1960s eliminated or truncated them severely once they realized that commercial applications would lie many years in the future. While not attaining the original vision of creating a truly thinking machine, research in artificial intelligence has generated numerous advances in expert systems, speech recognition, and image processing. Industry is actively commercializing many of these technologies and embedding them into a range of new products.

Virtual Reality

Innovation in computer graphics and virtual reality stems from the convergence of advances in numerous interrelated fields, such as computer graphics, psychology, computer networking, robotics, and computer hardware. It has been both pushed by technological advances in these underlying areas and pulled by creative attempts to devise particular applications, such as flight simulators, virtual surgery, engineering design, and tools for molecular modeling. Much of the underlying research has been conducted by universities, with federal support from agencies such as DARPA, the NSF, and NASA, but industry has played an important role in commercializing technologies and identifying key research needs. Interdisciplinary research efforts have been the norm in this field, as exemplified by the collaborative research effort between the computer graphics lab at the University of North Carolina, Chapel Hill, and Hewlett-Packard.

Networking
The nation's voice and data communications networks have different histories characterized by different relations between government and industry. Much of the infrastructure for voice communications was developed and deployed during a period in which AT&T enjoyed monopoly rights to the telephone market. This government-granted monopoly ensured widespread availability of service and effectively subsidized communications research. Subsequent development of data communications networking and the Internet grew largely out of government-sponsored research and deployment programs. DARPA funded development of packet switching as a collaborative effort with industry and academia. It subsequently created the interconnection protocols used over the Internet. The NSF provided additional funding for networking infrastructure for research and educational use and in effect laid the groundwork for today's Internet. The World Wide Web and browser technology currently used to navigate the Internet were devised by Timothy Berners-Lee at CERN and Marc Andreesen, then a student at the NSF-sponsored National Center for Supercomputing Applications at the University of Illinois.

Relational Databases
Development of relational database technology—now a billion-dollar industry dominated by such U.S. companies as Informix, Sybase, IBM, and Oracle—relied on the complementary efforts of industry and government-sponsored academics. Though originating within IBM, relational database technology was not rapidly commercialized because it competed with existing database products. The NSF funded the Ingres project at the University of California at Berkeley, which refined and promulgated the technology, thus spreading expertise and rekindling market interest in relational databases. Many of the companies now producing relational databases are populated by—or were founded by—participants in Ingres.

Theoretical Computer Science
Though typically viewed as the province of academia, theoretical computer science has benefited from the efforts of both industry and university researchers. While some advances—such as number theory and cryptology—have translated directly into practice, many others (such as finite state machines and complexity theory) have more subtly entered engineering practice and education, influencing the way researchers and product developers approach and think about problems. Progress in theory has both informed practice and been driven by practical developments that have challenged or outpaced existing theory.

Very-Large-Scale Integrated Circuits
Work on very-large-scale integrated (VLSI) circuits began in industry, with many companies devising proprietary design rules and forging only limited links to academic research. DARPA's VLSI program attempted to better link academic research to industry needs and to push the state-of-the-art, not only in semiconductor technology but in computer capabilities driven by such technologies. Research sponsored by DARPA at MIT, Stanford University, and the University of California at Berkeley resulted in several new architectures for parallel computing, reduced instruction set computing (RISC), and graphics (the geometry engine). Researchers from these programs assisted in commercializing the technology through start-up companies such as Thinking Machines, Sun Microsystems, and Silicon Graphics, respectively.

In the past, other forms of government support of technological change were common. During the first half of this century, the telephone industry flourished in the United States without substantial government funding but with government-granted natural monopolies. The patent system also provided means for industrial research laboratories, such as Bell Laboratories, to receive a return on their research and development investments. More recently, the government has supported the defining of technical standards, such as the Internet protocols, and standard programming languages, such as COBOL.

Federal procurement has also driven research and development. During the Semiautomatic Ground Environment (SAGE) project, the Air Force procured a number of advanced computers that were installed at MIT's Lincoln Laboratory. In the 1950s and early 1960s, many of the pioneers in computing learned through hands-on experimentation with these machines. The SAGE project can be compared to the learning experiences associated with the construction of the Erie Canal early in the last century. Contemporary engineers referred to the canal as the leading engineering school in the United States. Through grants placing computing equipment in engineering schools and universities, the NSF has also made possible hands-on learning experiences for countless young engineers and scientists. The DOE has also stimulated advances in supercomputers through procurement.

Besides diversity of funding, organizational innovation is a theme emerging from the history of computing and communications. In response to the insistence of Vannevar Bush, wartime head of the Office of Scientific Research and Development, and others that the country needed an organization to fund basic research, especially in the universities, Congress established the National Science Foundation in 1950. A few years earlier, the Navy founded the Office of Naval Research to draw on science and engineering resources in the universities. In the early 1950s, during an intense phase of the Cold War, the military services became the preeminent funders of computing and communications.

The Soviet Union's launching of *Sputnik* in 1957 caused concern in Congress and the country that the Soviets had forged ahead of the United States in advanced technology. In response, the U.S. Department of Defense, pressured by the Eisenhower administration, established the Advanced Research Projects Agency (ARPA, now DARPA) to fund technological projects with military implications. In 1962 DARPA created an Information Processing Techniques Office (IPTO), whose initial research agenda gave precedence to further development of computers for command and control systems.

With the passage of time, new organizations have emerged and old ones have often been re-formed or reinvented to respond to new national imperatives and to counter bureaucratic trends. DARPA's IPTO has transformed itself several times in order to bring greater coherence to its research efforts and to respond to technological developments. The NSF in 1986 formed the Computer and Information Sciences and Engineering Directorate (CISE) to couple and coordinate support for research, education, and infrastructure in computer science. The NSF, which customarily focused on basic research in universities, also began encouraging joint academic-industrial research centers. With the relative increase in venture capital and other private support of research and development in recent years, federal agencies such as the NSF have rationed their funding policies to complement funding by industry of short-term industrial research and development. Federal funding of long-term, high-risk initiatives continues to have a high priority.

As history suggests, federal funding agencies—established and yet to be established—will need to continue to adjust their strategies and tactics as national needs and imperatives change. The cold war imperative shaped technological history during much of the last half century. International competitiveness served as a driver of government funding of computing and communications during the late 1980s and early 1990s. With the end of the cold war and diminishing concerns about the competitiveness of the U.S. computer and communications industries, new missions may emerge as the rallying cry for technological development. Tomorrow, for instance, education or health may become the driving imperative.

Program Management

Individuals as well as organizations have shaped greatly the course of government funding over the past decades. The contributions of agency program managers are of critical importance but are not well known outside the managers' respective technical communities. Program managers in government funding agencies have responsibility for the initiation, funding, and oversight management of such projects as Project MAC and the ARPANET, which is the predecessor of the Internet. The most successful have married visions for technological progress with strong technical expertise and an understanding of the uncertainties of the research process. The funding and management styles of program managers flourished at ARPA during its early computer networking and artificial intelligence–funding decades. The activities of Joseph Carl Robnett Licklider provide a salient example of the program manager's role.

Head of ARPA's Information Processing Techniques Office and manager of projects from 1962 to 1964, Licklider came to ARPA on leave from the customary research and managerial activities at research universities and innovative computer firms. He was more familiar with the academic approach to problem solving and projects than the government's. After laying down extremely broad guidelines, Licklider preferred to draw specific project proposals from principal investigators or researchers in academic computer centers rather than define projects centrally. This style of funding and management allowed the government to stimulate innovation with a light touch, allowing researchers room to pursue new avenues of inquiry.

As further evidence of the light touch, government agencies, besides ARPA, manage the industrial and academic components of a funded system at an oversight level, leaving industry and universities considerable leeway in fulfilling contract specifications. In the case of grants, successful funding agencies often respond to agendas generated by researchers in the departments and centers. NSF grants have often supported virtually unfettered basic research that has produced significant advances.

Part of Licklider's success in using this style was his familiarity with leading research—and researchers—of the time. Working at the frontier of computer development, Licklider cultivated a small network of gifted researchers in the leading research universities. They and he had similar backgrounds, having mostly been educated or having taught in the Boston area, especially at MIT, and having worked with the early government-funded mainframe computers at MIT's Lincoln Laboratory.

As the field of computing has expanded, it has become more difficult for program managers to personally know an avant garde network of researchers and to intimately grasp details of diverse fields of inquiry.

Program managers now rely more on peer review and organizational procedures in deciding whom to support.

Why a Historical Approach?

This report and its lessons are grounded in a historical approach. By contrast, science and technology policy issues are usually approached in an analytical and quantitative way, which projects the future from the present by extrapolating quantitative data. A historical approach, as used in this report, assumes that the future may resemble the past as well as the present. Such a historical approach can provide a host of alternatives to current policy. For example, if another cold war involving the United States should break out, the role of government funding in sparking new technology might be more like the one played by the government in the 1950s than the one it plays today.

Furthermore, historical narrative accommodates messy complexity more easily than a tightly structured analytical essay. The approach also facilitates reflection on long-term process development and evolution. The case histories in this report present finely nuanced accounts, which also convey the ambiguities and contradictions common to real life experiences. An outstanding case in point is the SAGE project. Intended in the 1950s to provide a defense against air attack by bombers, SAGE's most influential and unintended long-term consequence was the training-by-doing of thousands of computer engineers, scientists, and software programmers. They subsequently staffed the nascent computing and communications revolution. The impact of the learning experience from this project was felt over the course of several decades.

Even though the historical approach offers insights, history cannot, however, demonstrate what might have happened if events had unfolded differently. For example, history can show the influence of federal funding on innovation in computing and communications, but it cannot suggest the direction the industry would have taken without federal intervention.

Acknowledgments

The author is indebted to Jerry Sheehan, senior program officer, NRC, and this study's director, for numerous suggestions. He is responsible for the preparation of the lessons and the summary of the case histories.

Reference

National Research Council, Committee on Innovations in Computing and Communications. (1999). *Lessons from history.* Washington, DC: National Academy Press.

Fax to Facts: Cold Fusion and the History of Science Information

Bruce V. Lewenstein

Abstract

Debates about the management of scientific information have traditionally assumed that information is generated first in scientific laboratories or field sites, becomes "real" or stable when it appears in peer-reviewed journals, and is then further disseminated through textbooks, encyclopedias, trade journals, government reports, mass media stories, and the like. Classic texts—from the proceedings of the 1948 Royal Society Scientific Information Conference and the 1958 International Conference on Scientific Information, through such modern texts as the annual proceedings of ASIS meetings—focus on how to classify and retrieve "real" scientific information, that is, how to retrieve laboratory and field reports that have been produced for and vetted by the peer review system. In this paper I use the cold fusion saga of the late 1980s and early 1990s to suggest that communication among scientists uses many more media than traditionally have been assumed. This particular historical episode suggests that we need to develop new models of the science information process, ones that account for permeable boundaries between formal publications, preprints, electronic computer networks, fax machines, mass media presentations, and other forums for scientific discussions.

Cold Fusion and the History of Science Information

Science has often served as the impetus for the analysis and improvement of information systems, perhaps symbolized best by Watson Davis. Trained as an engineer, for much of his career, Davis ran Science Service, a news bureau that provided science information to the public. He also helped create the international science fair system for youth. His experience in trying to stay on top of the burgeoning flow of specialized science information in the first half of this century led him to look for new ways of managing information, and he claimed to have coined the term *microfilm*. More important, he was a founder of the American Documentation Institute, forerunner of the American Society of Information Science (Lewenstein, 1988).

Another key figure in the general field of information science has been Eugene Garfield. While most scientists know him through the products of the company he founded and built, the Institute for Scientific Information (ISI), *information* scientists see beyond the specific ISI products to the fundamental insights into information management that he provided through his creation of citation indexing. The entire bibliometric field owes its origins to Garfield (Garfield, 1955, 1977–93).

Even today, as ASIS conferences devote special sessions to the challenges of managing information on the World Wide Web, such science-based topics as health provide most of the case studies.

But much of the work on information systems has drawn artificially sharp distinctions between primary science information—that is, original reports of specific research projects—and secondary information, such as media reports, textbooks, and government reports. (The Web is clearly an exception but a very recent one.) The key challenge in the field has often been seen as trying to serve both the producers and the users of primary information (who are often, of course, the same people). So, for example, if one looks up *science information* in the library, one finds lots of work on the management of peer-reviewed journals, preprint systems, and the like. Key founding texts in the field, such as reports of the Royal Society's Scientific Information Conference of 1948 and of the National Academy of Science's International Conference on Scientific Information a decade later, as well as NAS's 1969 report on scientific and technical communication, focus on the information use of primary scientific researchers. Key chapters have such titles as "Explorations on the Information Seeking Style of Researchers" or "Primary Communications," with sections on meetings, preprints, serials, and translations (National Academy of Sciences, 1959; Committee on Scientific and Technical Communication, 1969; Royal Society, 1948). Many of the theoretical models developed during this period, especially ideas about "invisible colleges" and reward systems and the like, focus on issues of information management within the world of primary scientific research (Crane, 1972; Hagstrom, 1965).

While that tradition of research has certainly reflected the reality of what most scientists mostly do on a day-to-day basis, it has presumed linear models of both science and communication. That is, the research has focused on the communication patterns within scientific research communities as if information is created there and then flows in a single direction, out to textbooks, industry, government, and the general public. Linear patterns have a long history in communication research; perhaps the best known is what we now call the "source-message-channel-receiver" model, first presented by the telephone engineer Claude Shannon and the mathematician Warren Weaver in 1948 (Shannon & Weaver, 1949). But today communication researchers consider such linear models to be outmoded. They suggest instead that we should focus on the interaction of multiple sources of information and on the way that meanings are shaped by the interactions.

In this paper I want to suggest that we need to reconceptualize scientific information systems in the same way; that is, we need to develop new models for science information systems that capture the complexity of communication interactions that shape science. To illustrate the need for new models, I will use the cold fusion saga that began (in a public way) in 1989.

The Cold Fusion Saga and Traditional Science Communication Models

The problem of relying on traditional models of science communication appears as soon as one tries to make sense of the cold fusion saga. From the moment of the initial press conference at the University of Utah on 23 March 1989, through the daily dispatches in newspapers around the world, to the widely quoted labeling of B. Stanley Pons and Martin Fleischmann as suffering from "incompetence and delusion," the mass media had a central place in the development of the science (Lewenstein, 1995b). Not only did the media inform the public about the development of a new area of scientific research, but for many scientists, the media also provided the forum for primary dissemination of technical information on a fast-moving research front. Unfortunately, traditional studies of science information provide little guidance for understanding how the media's presence in the debates affected the construction of cold fusion as a research area. Studies of the media's role have focused on issues of accuracy, balance, sensationalism, and relevance to the public. The inadequacies of this approach have been identified by a variety of researchers, who point to the essential similarities among all discourse that involves science. They also point to the assumption that science is only about "progress," only about a closer approximation to Truth, that underlies most analysis of science journalism; little research on science journalism looks at issues of trust, institutional authority, or other aspects of the social context of science (Dornan, 1988, 1990; Friedman, Dunwoody, & Rogers, 1986; Hilgartner, 1990; Krieghbaum, 1963, 1967; Nelkin, 1985; Shinn & Whitley, 1985).

The fundamental problem appears to be that traditional studies of science and the media are based on an outdated model of science information. During the 1960s, when sociologists and others developed the idea that "communication" is a fundamental part of science, the unidirectional, nonfeedback model of communication suggested by Shannon and Weaver was the most readily available theory (Berlo, 1960; Merton, 1973; Hagstrom, 1965; Cole & Cole, 1973; Ziman, 1968). It was in this context that thorough empirical studies of science communication were conducted from the late 1960s through the 1970s. Run by the psychologist William Garvey and a number of colleagues and informed by citation analysis, these studies provided a detailed description of the formal publication processes that scientific ideas go through as they move from the laboratory or blackboard into the realm of fixed and stable knowledge (Figure 1) (Nelson & Pollack, 1970; Garvey, 1979). Although Garvey and his colleagues did not cite the communication literature directly, the model they produced is clearly compatible with the linear dissemination-oriented SMCR model that had emerged in communication studies.

Beginning in the mid-1970s, sociologists of science began to react against the notion that scientific knowledge could be studied only as a privileged type of knowledge. Instead sociologists and anthropologists began to examine the everyday practices of scientists as they produced knowledge, and they questioned the idea that science is "created" in one sphere and then

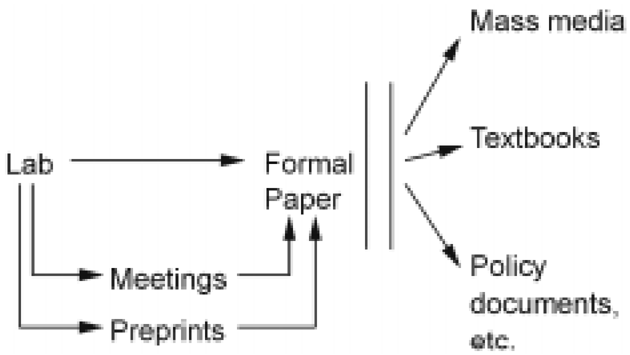


Figure 1. The formal communication system, as defined in the 1970s. Based on a diagram in William D. Garvey, Communication: The Essence of Science—Facilitating Information Exchange among Librarians, Scientists, Engineers and Students *(Oxford/New York: Pergamon Press, 1979), p. 169.*

disseminated in another, with distortion being an inevitable accompaniment of the dissemination. Instead researchers began to talk about expository science, emphasizing the way in which scientific information is shaped by the various audiences to which it can be addressed. At its core this new tradition argues that scientific knowledge does not exist in any abstract form but takes on shape and meaning only when it is expressed in specific contexts and addressed to specific audiences. According to this argument, a technical paper presented at a small workshop is no more "science" than is a multimedia extravaganza presented on an IMAX screen or at Disney World's EPCOT Center. Both are attempts to use rhetoric to present understandings of the natural world to particular audiences (Barnes, 1974; Bloor, 1976; Mulkay, 1979; Ravetz, 1971; Latour & Woolgar, 1979; Shinn & Whitley, 1985).

How does this newer view of the creation of science relate to cold fusion and the history of science information? It suggests that the reason analyses of cold fusion that look only at media coverage are unsatisfying is that they are based on an improper, or at least incomplete, understanding of the communication contexts in which the media reports appear. Thus, in what follows, I will provide a history of cold fusion that integrates media reports into the overall communication patterns that shaped the cold fusion saga. By doing so, I will show that the media's role in cold fusion can be understood only by reconceptualizing our models of science information flow.

The Public History of Cold Fusion

The public history of cold fusion began on 23 March 1989, when B. Stanley Pons and Martin Fleischmann announced at the University of Utah that they had found a way to produce nuclear fusion at room temperature in a small, relatively simple apparatus. Both the public and most other scientists first learned of Pons and Fleischmann's work through the mass media, by hearing breathless, excited reports on television and the radio (Cornell Cold Fusion Archive [CCFA], 1989a, March 23). Some scientists and members of the public had already read stories in the *Wall Street Journal* and *Financial Times* (of London), which both ran stories on the morning before the press conference (Bishop, 1989, March 23; CCFA, 1989c, March 23). The *Wall Street Journal*'s coverage was especially important because the next day it identified Steven Jones, the competitor at Brigham Young University who was doing work similar to Pons and Fleischmann's and whose activities were probably the stimulus that caused Pons and Fleischmann to go public when they did (Bishop, 1989, 24 March).

In the decade since that announcement cold fusion has gone through roughly four distinct periods (Figure 2) (Lewenstein & Baur, 1991; Lewenstein, 1992; Close, 1991; Huizenga, 1992; Mallove, 1991; Taubes, 1993). The first period, lasting about two months, appeared to many participants and observers as utter chaos (in the everyday, nonspecialist sense of that word). Claims and counterclaims changed almost daily; special cold fusion sessions were attached ad hoc to regular scientific meetings; stories with new and conflicting information appeared in newspapers, on the radio, on television, and on a newly created computer bulletin board. In the second period, through the summer and fall of 1989, much of the chaos disappeared, and the nature of the claims became clearer. Several special panels devoted to cold fusion issued reports; researchers identified topics of interest to them in the field; and for the most part public and scientific interest in the topic died off. The history since 1993 is less well covered, but can be followed on computer bulletin boards like the USENET newsgroup sci.physics.fusion and the Web sites http://www.mv.com/ipusers/zeropoint/ and http://world.std.com/~mica/cft.html. Eugene Mallove also publishes a cold fusion magazine, *Infinite Energy*, which contains much information on the continuing work by cold fusion believers.

In the third period, lasting throughout 1990, the sharp division between skeptics (or nonbelievers) and believers (as they were frequently labeled) became more prominent. On the first anniversary of the announcement the scientific journal *Nature*, home of the most prominent skeptics, published a scathing critical analysis of the situation in Pons and Fleischmann's own laboratory (Salamon et al., 1990). That same week believers gathered in Salt Lake City for the first Annual Cold Fusion Conference, sponsored by the Utah-funded National Cold Fusion Institute (Will, 1990). Later in the year the journal *Science* published a news article that came very close to accusing some cold fusion researchers of fraud (Taubes, 1990). In October, trying to tread a middle ground between belief and skepticism, Steven Jones organized a conference on "anomalous effects in deuterium/solid systems" at Brigham Young University; Pons and Fleischmann did not attend.

At the beginning of 1991 the division between skeptics and believers was vividly represented by the publication of two books on cold fusion with diametrically opposite evaluations of the state of the research field. Physicist Frank Close's *Too Hot to Handle* avoided accusing Pons and Fleischmann of fraud only by leaving open the possibility of sloppy incompetence, whereas Eugene Mallove's *Fire from Ice* predicted that cold fusion–powered home heaters were just around the corner. After that, in the fourth period, the two sides continued on their way, largely ignoring each other's critiques. Although a few skeptics (including Close, nuclear chemist John Huizenga, and nuclear physicist Douglas Morrison) continued to speak out against what they saw as the fraud and error of cold fusion supporters, most critics had long abandoned the field. Supporters, on the other hand, continued to meet: In addition to various regional gatherings, international meetings were held in Como, Italy (1991); Nagoya, Japan (1992); Maui, Hawaii (1993); Monte Carlo (1995); Hokkaido, Japan (1997); and Vancouver, British Columbia (1998). In essence, a new social group—a scientific subspecialty—had been created.

Communication and Chaos

In the first period chaos reigned. More accurately, information passed so quickly and permeably among multiple sources and multiple media that many participants recalled in interviews the sense of being completely inundated by information, without being able to judge the relative value of individual pieces of news or gossip. The interchangeability of media is particularly noticeable when we look at a basic information issue: how people heard about cold fusion. For example, then-MIT science writer Eugene Mallove, only a few months after the original announcement, could not recall whether he was in his office and his boss called him or he was out of the office and his boss told him when he checked in for the day (CCFA, 1989, November 8). Steve Koonin, a theoretical physicist at Caltech who was visiting Santa Barbara for a year, recalled who told him about cold fusion, but he did not remember whether the information came by electronic mail or telephone (CCFA, 1989, November 16). These confusions suggest that we need to be careful about focusing too closely on any one communication channel, without recognizing that users of those channels may not distinguish among them very carefully. Given the well-known phenomenon that people can recall precisely the circumstances in which they heard dramatic news, these examples may be anomalies. Other cold fusion participants recall with greater certainty how they heard of the new claims.

Another aspect of the complex flow of information was the degree to which various communication media began interacting within a day of the original press conference. In one example a science correspondent for National Public Radio used electronic mail to get interpretation of information that he had documented via audiotape. (The sci.physics bulletin board is one of thousands of bulletin boards available through the USENET

Cold Fusion Saga, 1989–1998

23 March 1989: *Public announcement*	
April–May 1989: *Media and scientific chaos*	
12 April	ACS/Dallas
26 April	U.S. Congress hearings
1 May	APS/Baltimore
23 May	Santa Fe conference
Summer–Fall 1989: *Growing stability*	
15 June	Harwell rejection
13 July	Interim DOE/ERAB report
15 October	NSF/EPRI panel
12 November	Final DOE/ERAB report
1990: *Consolidation of positions*	
29 March	1st NCFI CF Conference
15 June	*Science* charges fraud
22 October	BYU conference on anomalous effects
1991–1998: *Ongoing work, two separate strands*	
January 1991	Pons resigns from University of Utah
Spring 1991	Close & Mallove books
June 1991	2nd CF conference, Italy
January 1992	Riley killed at SRI
October 1992	3rd CF conference, Japan
December 1993	4th CF conference, Hawaii
April 1995	5th CF conference, Monaco
Fall 1996	6th CF conference, Japan
Spring 1998	7th CF conference, Vancouver

Figure 2. Major points in the cold fusion saga timeline.

computer network, a worldwide collection of "newsgroups" used in the early 1990s by at least 1.4 million people. The Internet has, of course, grown rapidly since then.) For other observers videotape was more important; researchers phoned the University of Utah, requesting copies of a videotape showing the press conference, or watched copies of the television shows that had run extensive stories on the announcement, despite the fact that the level of detail in these programs was not high (CCFA, 1989b, March 23). Some researchers turned to the actual press release for more information, but they did not find much: "In the experiment, electrochemical techniques are used to fuse some of the components of heavy water, which contains deuterium and occurs naturally in sea water" (CCFA, Press Release, Fogle folder, 1989).

Instead researchers found themselves turning the mass media into a source for technical data: "We used photographs from the *L[os] A[ngeles] Times* of Pons holding the cell, and you could see pretty well how it was made," said Michael Sailor, a Caltech postdoctoral student in electrochemistry. "We used Pons's finger for a scale. Gordon [Miskelly, another postdoc] figured his hand was about equal-sized, so he scaled it to his own finger." Another Caltech student brought in the videotapes. "We looked at them to find out what the readings on their thermistors were, where the electrodes were, and how they were doing their electrochemistry," said Nathan Lewis, professor of electrochemistry at Caltech (Smith, 1989).

The traditional models of science communication, by focusing on peer-reviewed publications, assume that scientists work with stable, certain information. But the cold fusion saga, like so many controversies, opens up the inner workings of science and lets us see the daily workings of science in greater detail. As in any fast-moving area of science, researchers lacked access to a fixed, stable piece of information (a preprint or published article); so many scientists began exchanging rumors, newspaper articles, and so on. Faxed copies of newspaper articles from distant countries, accompanied by handwritten comments on the article or on other developments, soon circulated widely (CCFA, Manos folder). The combination of newspaper, fax, and interpersonal communication all shaped the meaning of any one particular piece of information. Attempting to sort out the impact of each component would do injustice to the complex context of communication.

After the first week scientists and reporters began to receive preprints and then reprints of various technical articles (CCFA, Preprints folders). Not only did researchers need to acquire, read, and process the information in each of these texts, they also had to compare them—especially the differences between the early manuscripts and the final published articles. Although the process of sorting out the differences and making judgments about the multiple texts would eventually lead to greater stability of information, many researchers recalled in interviews that the need to first resolve *which* version someone was talking about contributed to the sense of chaos or instability.

An important issue concerning access soon emerged: Different people had different levels of access to information. By the time preprints and publications began to get wide circulation in late April, some people had had access to them for almost a month. For example, Richard Garwin, a physicist at IBM, had been asked by *Nature* to referee both a manuscript from Fleischmann and Pons and the Jones manuscript around the first of the month. In late April, Garwin's own summary of a one-day cold fusion conference in Erice, Sicily, appeared in *Nature*, concluding that "large heat release from fusion at room temperature would be a multidimensional revolution. I bet against its confirmation." But while this summary appears, in the text, to be based solely on the presentations at Erice, and while Garwin was careful not to cite his privileged access to the original manuscripts, the extra several weeks he had to consider information undoubtedly shaped his analysis. (In addition, of course, Garwin [1989] knew that the information to which he had access was direct from the main protagonists rather than filtered through mass media reports or other communication media.)

To understand the importance of Garwin's privileged access, recall that his article was one of the first to reach print. Not until the following issue of *Nature* was Jones's article published, along with commentaries by several other scientists. Readers no doubt made some judgments about the relative importance of information in speculative letters, Garwin's meeting report, and Jones's complete article. Communication theory suggests, however, that those judgments are extremely complex and not likely to be directly related to "objective" measures of the relative importance assigned to each publication. People take in lots of information, filter it in various ways, and base their judgments on a range of issues running from salience and importance through time of day and state of hunger. In the case of cold fusion readers had to judge the value of suggestions published by prominent scientists (Nobel laureate Linus Pauling published a letter early on, for example) versus letters from physicists and chemists in Utah (who, to outsiders, might be presumed to have more detailed local knowledge). Theory suggests that each reader would make a different judgment, based on completely contingent factors. No model attempting to predict the value of different types of communications works (Bryant & Street, 1988; Dervin, 1989).

Another factor that made it more difficult for researchers and others trying to make stable judgments about cold fusion was the presence of new or unusual patterns of information flow. Some members of the media, for example, agreed to serve as brokers in the information exchange among scientists. Those activities went beyond merely passing around copies of papers and negotiating access to information. Sometimes reporters acted explicitly as mediators among scientific sources. David Ansley, a reporter for the *San Jose Mercury News*, recalled that:

> At one point, I called up [University of Utah vice president for research James] Brophy and said "Look, this is making no sense. You say that all it takes is the simple description and that other researchers ought to be able to duplicate it. . . . [But] here are the questions they're asking me. Can you answer any of these questions?" And he would give me the answers. I would call [the researchers] back, and they would say "That's so simplistic. That's just not enough. We need X, Y, and Z. The way he's describing that doesn't do us any good." I'd call [Brophy] back, and he'd say, "No, really, that's how it works. It's that simple." (CCFA, 1989, November 18; see also CCFA 1989, August 11 & July 12)

For those people following the rapidly expanding electronic bulletin boards, the mix of media also applied. By the beginning of April a separate newsgroup, completely devoted to cold fusion, called "alt.fusion" was created. Early messages ranged from personal summaries of a seminar given by Fleischmann at CERN, to brief snippets announcing that "CBS News is reporting that the Pons-Fleischmann experiment has been reproduced in Hungary," to speculations about the potential impact of cold fusion on oil prices and the world economy (CCFA, e-mail file).

Thus, no matter where researchers and others trying to find stable information turned to stay informed, the barrage of conflicting material about cold fusion led to what the media frequently called "fusion confusion." The sense of instability caused by frequently changing judgments was reflected in newspaper coverage. At the *Los Angeles Times*, experienced science writer Lee Dye wrote on 19 April that Pons and Fleischmann were receiving a "flood of support"; two days later he said that "evidence continued to mount in support of the controversial experiment." Yet just two days after that, on 23 April, he began a story by noting that "scientists at major research institutions throughout the country are growing increasingly frustrated over their inability to replicate a supposedly simple experiment" (Dye, 1989, 19, 21, 23 April).

To get a sense of the instability, consider what might have happened over just two days. On the evening of Monday, 1 May 1989, a parade of speakers at the American Physical Society meeting in Baltimore ridiculed cold fusion. Strong critiques were made of various experiments from which scientists had claimed positive results. Theoretical calculations were presented to show that Fleischmann and Pons's claims violated the predictions of nuclear theory by nearly forty orders of magnitude. At a press conference eight of nine researchers voted against the likelihood that cold fusion would prove to exist. The sense that Fleischmann and Pons had made absolutely elementary mistakes and that cold fusion could be rejected out of hand was captured by one physicist who wrote a piece of doggerel to criticize the temperature measurements of a colleague:

> Tens of millions of dollars at stake, dear brother,
> Because some scientists put a thermometer
> At one place and not another.

And Caltech's Koonin was widely quoted when he said that "we are suffering from the incompetence and possible delusion of Professors Pons and Fleischmann" (Associated Press, 1989; Browne, 1989; CCFA, audiotapes and videotapes; CCFA, 1989, May 22).

On the following day, Tuesday, 2 May, MIT researchers led by Richard Petrasso submitted to *Nature* a major article questioning the gamma-ray spectrums presented by Fleischmann and Pons as evidence of nuclear reaction products. (Petrasso's article included a gamma-ray spectrum taken off a television broadcast; this may be the first time a piece of scientific evidence has carried a citation to "KSL-TV in Utah." This unusual reference highlights interactions between media that information analysts have not previously noticed [Petrasso, Chen, Wenzel, Parker, Li, & Fiore, 1989].)

Yet that week *Time* and *Newsweek* issued their 8 May 1989 magazines. Both chose to feature cold fusion on the cover. Though the headlines included some skepticism (*Time*'s was "Fusion or Illusion: How Two Obscure Chemists Stirred Excitement—and Outrage—in the Scientific World"), the effect was to present cold fusion as a potential energy savior to millions of people around the world. A reader had to contrast the weekly news magazines, which by their writing style foster a sense of authoritativeness, with the reports of the APS meeting appearing in their daily newspapers. Especially for readers who depended on brief stories in local papers or television broadcasts, the news magazine stories might well have had more impact. And so, while journalists and researchers who had attended the APS meeting decided that consensus—or a stable judgment—was becoming clear, researchers not physically present in Baltimore, and certainly the general public, still faced highly unstable information.

The period of instability ran through the end of May, when the Department of Energy sponsored a three-day meeting devoted to cold fusion in Santa Fe, New Mexico. By the end of the meeting many of the four hundred participants were still undecided about the reality of cold fusion effects, but they were much clearer about how to go about testing the claims of Pons and Fleischmann, Jones, and the others who had now entered the fray. As *Science* magazine said in its headline, it was the "End of Act I" (Pool, 1989).

The Growth of Stability and Consolidation

Although the cold fusion drama continued after the intermission that (metaphorically) followed the Santa Fe meeting, the mass media for the most part did not come back to the show. The peak of media coverage of cold fusion occurred during the excitement of the mid-April period (Figure 3), when fresh reports appeared daily and Pons was cheered by seven thousand chemists at the American Chemical Society meeting in Dallas. A dramatic drop in coverage came after the APS meeting in Baltimore; many reporters said in interviews that the apparent consensus among scientists meant that a stable judgment had appeared and they could turn their attention to new issues. And following the Santa Fe meeting coverage dropped even more (Lewenstein, 1992; University Microfilms, Inc.).

With the drop in media coverage, the number of communication channels involved in cold fusion dropped dramatically. Without the mass media to carry information from one channel to another, the intermixing of other communication media also dropped, suggesting that, as we think about a more complex model of science communication, we need to give the mass media a catalyzing role in creating complexity.

The drop in media coverage, however, does *not* imply that cold fusion itself died out after May 1989. Indeed, there is significant evidence to show that cold fusion research remained robust for months after the Santa Fe meeting, even among the harshest skeptics. Reports of the Santa Fe meeting were circulated by electronic mail, then printed out and circulated even further on paper (CCFA, Weisz folder). The Department of Energy had created a special panel to investigate cold fusion. That panel met for the first time at the Santa Fe meeting, then conducted a series of meetings and site visits over the summer. When the Energy Research Advisory Board (ERAB), as the DOE panel was known, issued an interim report in mid-July, press coverage labeled the report a devastating blow to cold fusion. And while the report certainly was not friendly to cold fusion, it explicitly acknowledged the need for further research (CCFA, 1989, July 13).

The ERAB panel's report was part of the emerging consensus during the summer. About the same time a Brookhaven National Laboratory researcher presented a paper titled "Cold Fusion: Myth or Reality?" He concluded that "Cold fusion will *not* be our next power source," but that "there *do* appear to be some interesting physical effects to be pursued" (CCFA, Brookhaven National Laboratory). In the meantime the state of Utah had allocated about $5 million to a new National Cold Fusion Institute in the University of Utah's research park, and experiments there were being conducted with the advice of Pons and Fleischmann.

During the fall continuing discussions among the many participants took place at meetings and via the traditional forms of scientific communication, especially preprints and papers. A two-day cold fusion meeting was held at Varenna, Italy. The National Science Foundation and the Electric Power Research Institute (funded by the electric utility industry) jointly sponsored a three-day meeting in Washington in October. The ERAB panel issued its final negative report in November. But the results of these various meetings and panels were also distributed electronically and via fax and telephone into a growing cold fusion underground. Douglas Morrison, a CERN physicist who was one of the first and most persistent to tag cold fusion as pathological science, distributed an irregular *Cold Fusion Newsletter* via electronic mail, and copies were posted to the sci.physics.fusion newsgroup (which had, by now, superseded the alt.fusion newsgroup) as well (CCFA, Morrison newsletter folder).

Thus, by the end of 1989, the cold fusion saga had become stable. Mass media coverage of cold fusion (including news reports in the science trade press, such as the news sections of *Nature* and *Science*) dropped essentially to zero by the fall and remained there except for brief flurries caused by anniversaries of the original announcement or by accusations of fraud that have periodically appeared. Meanwhile the number of articles appearing in the technical refereed literature had climbed steadily and by the end of 1989 consistently averaged nearly twenty articles per month. Electronic newsgroup volume was also about to settle into a pattern and by early 1990 averaged about seventy messages per month.

Another island of stability grew out of the efforts of some researchers who deliberately removed themselves from the morass of information in which they found themselves wallowing. David Williams, an electrochemist who led the replication effort at the United Kingdom's Harwell laboratory, had begun his experiments with help from Fleischmann *before* the public announcement. After the announcement, he briefly noted the many conflicting bits of information he heard from other groups attempting replications. Recognizing the confusion this was creating in his own group's work, he made a conscious decision to disregard information coming from outside Harwell. His group felt that they should focus on their own experiments rather than trying to follow every twist and turn that others reported (CCFA, 1990, April 11). Charles Martin, an electrochemist at Texas A&M University who had been among the first apparently to replicate parts of the Pons and Fleischmann experiments, discovered in the early summer of 1989 that he had devoted so much time to cold fusion that he had dropped all other activities—including keeping up his log book and playing racquetball. He, too, made a conscious decision to resume his normal information and working habits—which, of necessity, meant spending less time seeking information and watching for the latest permutations in the cold fusion activities of others (CCFA, 1989, July 17).

As 1990 proceeded, the stable positions consolidated. Review articles and conference proceedings that argued for cold fusion began to appear, such as an Indian summary of one hundred experiments performed at the Bhabha Atomic Research Center in Trombay, Bombay, and the proceedings of the First Annual Conference on Cold Fusion, sponsored in March 1990 by the NCFI (Iyengar & Srinivasan, 1989; Will et al., 1990). Notice that much of this information continued to appear in the "gray literature"—accessible to insiders who were on distribution lists, but not part of the formal peer-reviewed literature system. Out of the NCFI conference came comments indicating the strength of the beliefs of cold fusion supporters: "It is no longer possible to lightly dismiss the reality of cold fusion," said UCLA physicist Julian Schwinger, a 1965 Nobel laureate. Recent calorimetric results "will be noted as a decisive turning point in the history of the affair," said Ernest

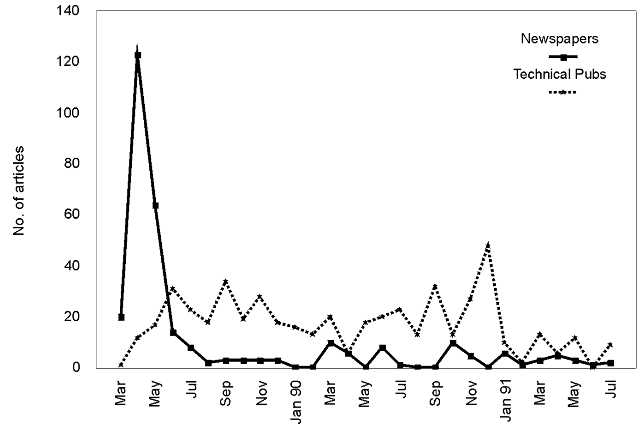


Figure 3. Cold fusion publications. Newspaper data (which include book reviews) from the "Newspaper Abstracts OnDisc" CD-ROM database. Technical publications taken from the Cold Fusion Bibliography distributed via sci.physics.fusion Internet newsgroup by chemist Dieter Britz; data shown include only those items for which the specific month of publication is identified.

Yeager, a Case Western Reserve researcher. "These results cannot be explained by trivial mathematical errors," Yeager continued. And two Los Alamos National Laboratory researchers, speaking to one of the specific issues that bothered many observers, said, "We can put aside the question as to whether the tritium is real." To their satisfaction, it was (Mallove, 1991).

Finally, in mid-1990, Fleischmann and Pons published the major article they had been promising for months, providing in exhaustive detail the calculations they had performed to calculate the excess heat they said they had observed in their cells (Fleischmann, Pons, Anderson, Li, & Hawkins, 1990).

The skeptics, however, were also consolidating their position, and the new article from Fleischmann and Pons contributed to their certainty, since it dealt only with calorimetry, not with measurements of nuclear reaction products. The skeptics pointed to the lack of evidence of nuclear reactions to justify their own decision to ignore further cold fusion claims. They were especially impressed by a paper published in *Nature* on the first anniversary of the original announcement by Michael Salamon, a University of Utah physicist who had been allowed into Pons's laboratory and had found no evidence of nuclear reaction products (Salamon, et al., 1990).

Skeptics could also point to the gradual decrease in the number of publications in the formal refereed literature. The decrease was especially dramatic if one considered the actual date of submission, rather than the date of publication. Submission dates showed that the bulk of published papers actually represented research done in 1989. The volume of research conducted after that was clearly dropping (Lewenstein, 1992).

The split between skeptics and believers was perhaps best illustrated by the publication in early 1991 of two books: physicist Frank Close's *Too Hot to Handle*, an indictment of the methods and procedures followed by Pons and Fleischmann; and science writer Eugene Mallove's *Fire from Ice*, a paean to the possibilities of power created by cold fusion.

After 1991 cold fusion was essentially completely divided into the two paths of belief and skepticism, with few intersections between them. Although a few traditional journals continued to publish cold fusion work (most notably *Fusion Technology*), communication now tended to take place between individuals, in informal meetings or via the "cold fusion underground" of telephone and fax communications. The proceedings of the annual cold fusion meetings were also important sources of information for continuing cold fusion researchers, as were newsletters like *Fusion Facts* (published in Salt Lake City) and magazine's like Mallove's *Infinite Energy*.

Electronic conversations about cold fusion continued to take place regularly in the sci.physics.fusion newsgroup and the associated Fusion Digest listserv distributed over the Internet. Until about mid-1992 the newsgroup consisted primarily of interested bystanders commenting on cold fusion. But with the regular contributions of a few active cold fusion researchers or supporters, volume increased somewhat after that. (The growth may have reflected new developments within the cold fusion social community as well as the rapid growth of all Internet-based activities worldwide; exploring those developments, however, is beyond the scope of this article.)

The mass media continued to run an occasional story on cold fusion. But for the most part the complexity of the cold fusion communication context had died out by the end of 1992.

Conclusion

Two major conclusions can be drawn from this history of cold fusion focusing on communication issues.

Communication Complexity

Although traditional models of science communication described a linear process, this article has clearly shown that many forms of scientific communication interacted in the case of cold fusion. A better, nonlinear model might be a circle or a sphere, with all forms of communication leading to each other (Figure 4). Some evidence of mixed forms of communication makes this clear:

- The reliance of some teams on television for depicting experiments that they tried to reproduce.
- The debate on social and moral issues (such as the effect of cold fusion on the world economy) appearing almost solely on the electronic networks, but drawing from data mainly in the mass media.
- The exchange of information among media, such as the NPR reporter who gathered commentary on the Internet or the media commentary that appeared on the Internet.
- The growing sense of excitement after the Jones preprint was distributed via fax and electronic mail, with the excitement infecting the mass media.
- The importance of meetings, both large and small, for setting the tone among multiple media.
- The way in which some researchers changed their opinion of Pons and Fleischmann (generally in the negative direction) after they appeared before a congressional hearing on 26 April and tapes of their appearance were broadcast on C-SPAN.

In this model, the category "mass media" moves toward a central place. As suggested in the text, mass media were not crucial to the ongoing process of cold fusion science. But their presence did contribute to the complexity and instability of information available to researchers at any given time. The mix of all communication media depended on the degree to which *mass* media were involved.

This revised model of the science communication process suggests a resolution to one of our initial problems: how to understand the role of the mass media in science. The answer is do not try—or, at least, do not try without also examining the full communication context. In the cold fusion saga any attempt to understand the role of the mass media must deal with the permeable boundaries that existed between the various forms of communication that were involved. In more general terms the model suggests that to understand science communication, we must explore the complexity of interactions among *all* media.

One can question whether this more complex version of science communication applies to all of science. In the science studies world research on scientific controversies is valued precisely because it highlights points of stress in the system. By that argument a model derived from studies of cold fusion is a plausible candidate for explaining the communication patterns seen in other areas of science. But the role of public discussion of fine details of the scientific process was clearly greater in the cold fusion saga than in most areas of science; conceivably this could bias my description toward greater complexity than normal. Only future studies attempting to apply this description of science communication can resolve this issue.

Preliminary evidence suggests, however, that mass media does indeed influence scientific practice. For example, in a study of research patterns appearing in the *New England Journal of Medicine*, sociologist David Phillips and colleagues showed that those articles that had been brought to the attention of the public by the *New York Times* received an amplified response in the technical scientific literature for years after they initially appeared. Several analyses of recent controversies in geology regarding catastrophes and extinctions have pointed to the media's role in catalyzing technical discussions. A brief analysis of the media's role in the discussion of the possibility that fossil signs of life were found in a Martian meteorite also supports the importance of understanding the interactions of media (Phillips, Kanter, Bednarczyk, & Tastad, 1991; Clemens, 1986; Glen 1994; Lewenstein, 1997).

In perhaps the closest comparison to cold fusion a high degree of complexity occurred in the case of high-temperature superconductivity. In the early months of that field scientists regularly presented data straight out of the laboratory at press conferences and other nontraditional forums. As in the case of cold fusion, researchers from other laboratories had to decide whether to wait for more stable, certain information or to proceed with their own work based on the incomplete information acquired through the media. The media played a role in helping researchers exchange data, though with unclear results on the progress of the research itself (Hazen, 1988; Schechter, 1989; Felt, 1993; Nowotny & Felt, 1997). Superconductivity represents the opposite pole from cold fusion: an unexpected finding that eventually led to the consensus that the phenomenon had been confirmed. Yet it also offers a case in which the model described above seems, to a first approximation, to be applicable.

Despite these suggestive cases, more work is needed to see if the model of complex science communication described above can be applied in other contexts.

Information Stability

One of the most intriguing new questions in information science is the effect of new communication technologies on the process of scientific inquiry (Lewenstein & Heinz, 1992; Harrison & Stephen, 1994; Crawford, Hurd & Weller, 1996). This study suggests that one important issue is the degree to which scientific judgments are based on the stability of information. Cold fusion presents a particularly vivid example of the ways in which judgments changed depending on what information was available. Clearly the nontraditional forms of communication (including electronic mail, electronic bulletin boards, faxes, and news media reporting) were associated with unstable information. But what was their role? Did the presence of new communication contexts create instability? Or were the new contexts—and the vast quantities of material they offered—used precisely because they provided an opportunity to resolve uncertainty and thus create stability more quickly than traditional contexts? There is a correlation, but in what direction is the causation: Does information cause instability, or does instability create a need for information? Is it even possible, given the interactional model of science communication presented above, to specify direction or causality?

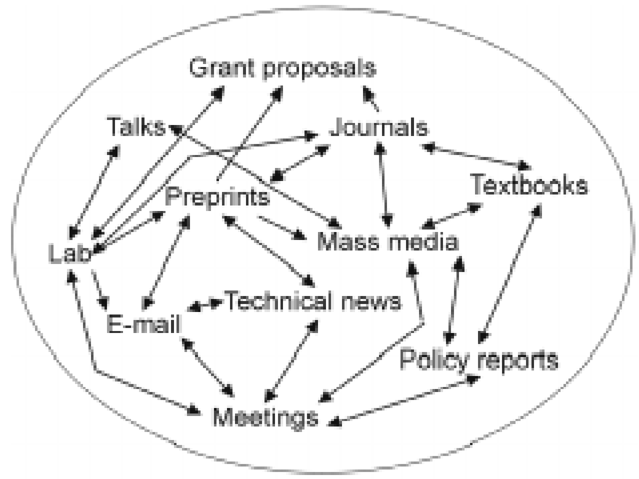


Figure 4. The web of science communication contexts.

Although there is not yet sufficient clear evidence to answer these questions, I want to present one possible answer, in part to stimulate further discussion. I believe the available evidence suggests that, in the cold fusion case, new communication contexts (including electronic technologies) ensured a surfeit of information; that this surfeit led to confusion and complexity; and that only when the mass media dropped out of the communication context did the scientific community proceed to more stable information and more stable judgments (both among skeptics and believers). At the same time I think that the initial presence of complex, unstable information also created the need to find stability more quickly and thus may have hastened the time when

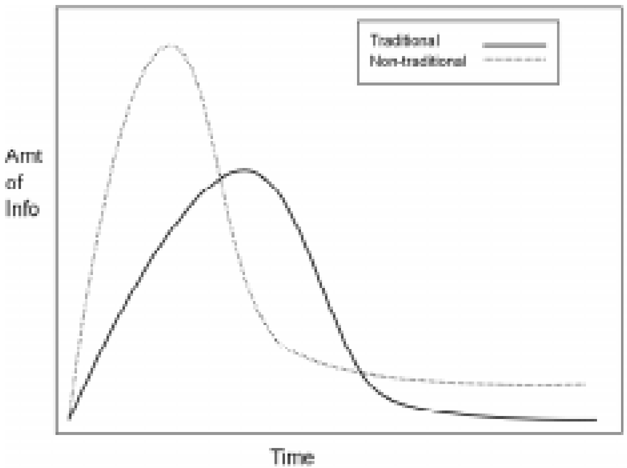


Figure 5. Information stability. The horizontal axis represents time; the vertical axis represents information quantity. The traditional curve, with less information spread out over a greater amount of time, is lower and more "stable." The new curve, with greater information reached initially, shows a thinner, less stable "peak." Notice that the new curve also levels out at a higher level, suggesting that information in nontraditional contexts remains more complex over time, despite reaching a relatively stable level earlier.

stable judgments were formed. More information led to more instability but also reduced the time until stability was achieved. This relationship is illustrated in Figure 5; while the figure can only be suggestive, since it lacks units, it may provide a useful graphical metaphor as we try to develop new models of science communication.

Although the instantaneous nature of modern electronic communication has become a cliché, the speed with which information flowed had an important impact on cold fusion, because many people were trying to make decisions based on a mish-mash of changing data, of varying degrees of reliability, and in various states of intelligent presentation. As the model presented in Figure 4 suggests, information flow in science is a convoluted, irregular process. The pressure of e-mail and other forms of electronic communication (in addition to the presence of the mass media) *added* to the confusion in the cold fusion case. Communication times were shorter, but the communication itself was more complex, chaotic, and intense. Only after information channels were removed, and thus the chances of receiving conflicting or competing information reduced, could stability develop.

What might be the effects of a shorter, more intense communication period in which more unstable information is converted into stable knowledge? Two possibilities exist, which need to be investigated with additional research:

1. Greater complexity could change the way in which people are recruited into the scientific debates, since it changes the premium placed on access to information, speed of response, etc. (Some people, for example, have argued that electronic mail allows the scientific playing field to become more level, since issues of status, age, gender, physical location, and so on do not enter into an electronically mediated discussion in traditional ways. But at least in the cold fusion case, it is not clear that such democratization happened [Lewenstein, 1995a].)
2. Another possible effect is that intense communication periods may make emotion more important: Anyone who uses e-mail regularly has had the sensation of pushing the SEND button and then saying, "Oops, I didn't really want to say that." With an old-fashioned letter, or a game of telephone tag before you reach someone, there is the chance for things to cool down a bit. Emotion, of course, plays no role in the canonical "scientific method." But given the clear findings of science-studies researchers regarding the importance of social interaction in the development of scientific knowledge, we need more research on the role of emotion in scientific communication (LaFollette, 1990).

Although we do not fully understand these effects, one possibility is that the traditional routines of peer review and formal publication will remain important components of the social process of science, because they will serve as ways for information to become more stable than it is in the faster but more ephemeral forms of communication that are a part of everyday scientific life. As the density of communication media falls off, information (and thus knowledge?) becomes more stable because the competing sources of information are not there.

Clearly these possibilities are only speculation, constrained by our lack of clear knowledge of how the science communication process actually works. While the traditional linear models focusing on peer-reviewed literature have provided useful guides for much of the last generation, they are inadequate to explain the complexity of modern scientific communication. We must develop more sophisticated models of science information, both for theoretical reasons and as a guide to the practice of librarians, information scientists, and scientific researchers in the future.

Acknowledgments

Much of this paper originally appeared as Bruce V. Lewenstein, (1995), From Fax to Facts: Communication in the Cold Fusion Saga, *Social Studies of Science, 25*(3), 403–436. Research was supported by grants from the National Science Foundation; the U.S. Department of Agriculture; the Chemical Heritage Foundation; and the Department of Communication, the Department of Science and Technology Studies, the Division of Rare and Manuscript Collections, and the Dean of the College of Agriculture and Life Sciences at Cornell University.

References

Associated Press. (1989, May 3). More scientists cool to fusion claims. *San Jose Mercury News.* p. 7A.

Barnes, B. (1974). *Scientific knowledge and sociological theory.* London: Routledge & Kegan Paul.

Berlo, D. (1960). *The process of communication.* New York: Holt, Rinehart and Winston.

Bishop, J. E. (1989, March 23). Breakthrough in fusion may be announced. *Wall Street Journal.* p. B1.

Bishop, J. E. (1989, March 24). Utah scientists claim breakthrough in quest for fusion energy. *Wall Street Journal.* p. A1.

Bloor, D. (1976). *Knowledge and social imagery.* London: Routledge & Kegan Paul.

Browne, M. W. (1989, May 3). Fusion claim is greeted with scorn by physicists. *New York Times.* p. A1.

Bryant, J. & Street, R. L., Jr., (1988). From reactivity to activity and action: An evolving concept and Weltanschauung in mass and interpersonal communication. In R. P. Hawkins, J. M. Wiemann, & S. Pingree (Eds.). *Advancing communication science: Merging mass and interpersonal process,* (pp. 162–190).Newbury Park, CA: Sage.

Clemens, E. S. (1986). Of asteroids and dinosaurs: The role of the press in the shaping of scientific debate. *Social Studies of Science*, *16*, 421–456.

Close, F. (1991). *Too hot to handle: The race for cold fusion.* London/Princeton: W. H. Allen/Princeton University Press.

Cole, J., & Cole, S. (1973). *Social stratification in science*. Chicago: University of Chicago Press.

Committee on Scientific and Technical Communication, National Academy of Sciences—National Academy of Engineering. (1969). *Scientific and technical communication: A pressing national problem and recommendations for its solution.* Washington, DC: National Academy of Sciences.

Cornell Cold Fusion Archive, collection #4451, Division of Rare and Manuscript Collections, Kroch Library, Cornell University (CCFA). (1989, July 12). Kelvin Lynn taped interview.

CCFA. (1989, August 11). Ivan Amato taped interview.

CCFA. (1989, July 13). Typescript. Steven Jones folder.

CCFA. (1989, November 16). Steven Koonin taped interview.

CCFA. (1989, July 17). Charles Martin taped interview.

CCFA. (1989, November 18). David Ansley taped interview.

CCFA. (1989, May 22). Phillip Schewe to William Havens. James Krumhansl [Part 1] folder.

CCFA. (1989, November 8) Eugene Mallove taped interview.

CCFA. (1989a, March 23). *CBS Evening News.* Vanderbilt TV Archives videotape.

CCFA. (1989b, March 23). Off the Record videotape.

CCFA. (1989c, March 23). "Simple experiment" results in sustained n-fusion at room temperature for first time. University of Utah press release. Pamela Fogle folder.

CCFA. (1990, April 11). David Williams taped interview.

CCFA. APS press conference audiotapes.

CCFA. APS, TV news excerpts videotapes.

CCFA. Brookhaven National Laboratory. Seminar notes folder.

CCFA. Dennis Manos folder.

CCFA. Diane Weisz folder.

CCFA. E-mail file. Messages from John Gilmore and Matt Ginsberg. 890405.af electronic file.

CCFA. Morrison newsletter folder.

CCFA. Preprints folders.

CCFA. Richard Garwin folder.

Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities.* Chicago: Chicago Press.

Crawford, S. Y., Hurd, J. M., & Weller, A. C. (1996). *From print to electronic: The transformation of scientific communication*. Medford, NJ: Information Today.

Dervin, B. (1989). Audience as listener and learner, teacher and confidante: The sense making approach. In R. E. Rice & C. K. Atkin (Eds.). *Public communication campaigns* (2nd ed.), (pp. 67–86). Newbury Park, CA: Sage.

Dornan, C. (1988). The "problem" of science and the media: A few seminal texts in their context, 1956-1965. *Journal of Communication Inquiry*, *12*, 53–70.

Dornan, C. (1990). Some problems in conceptualizing the issue of "science and the media." *Critical Studies in Mass Communication*, *7,* 48–71.

Dye, L. (1989, April 19). Stanford lab adds support to table-top fusion claims. *Los Angeles Times.* pp. 3, 17.

Dye, L. (1989, April 21). Fusion scientists delay publication of research. *Los Angeles Times.* pp. 3, 28.

Dye, L. (1989, April 23). Scientists grow frustrated over fusion claim. *Los Angeles Times.* pp. 1, 43.

Felt, U. (1993). Fabricating scientific success stories. *Public Understanding of Science*, *2,* 375–390.

Fleischmann, M., Pons, S., Anderson, M. W., Li, L. J., Hawkins, M. (1990, July 25). Calorimetry of the palladium-deuterium-heavy water system. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, *287,* 293–348.

Friedman, S. M., Dunwoody, S., & Rogers C. L. (Eds.). (1986). *Scientists and journalists: Reporting science as news.* New York: Free Press.

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science, 122,* 108–111.

Garfield, E. (1977–1993). *Essays of an Information Scientist* (Vols. 1–15). Philadelphia: Institute for Scientific Information.

Garvey, W. D. (1979). *Communication: The essence of science. Facilitating information exchange among librarians, scientists, engineers and students.* Oxford/New York: Pergamon Press.

Garwin, R. (1989, April 20). Consensus on cold fusion still elusive. *Nature*, *338*, 616–617.

Glen, W. (Ed.). (1994). *The mass extinction debates: How science works in a crisis*. Stanford, CA: Stanford University Press.

Hagstrom, W. (1965). *The scientific community*. New York: Basic Books.

Harrison, T. M., & Stephen, T. D. (Eds.). (1994). *Computer networking and scholarship in the 21st century university*. Albany, NY: State University of New York Press.

Hazen, R. M. (1988). *The breakthrough*. New York: Summit Books.

Hilgartner, S. (1990). The dominant view of popularization: Conceptual problems, political uses. *Social Studies of Science*, *20*, 519–539.

Huizenga, J. (1992). *Cold fusion: The scientific fiasco of the century.* Rochester, NY: Rochester University Press.

Iyengar, P. K., & Srinivasan, M. (Eds.). (1989). *BARC studies in cold fusion.* Trombay, Bombay: Bhabha Atomic Research Centre (copy in CCFA, Papers folder).

Krieghbaum, H. (1967). *Science and the mass media.* New York: New York University Press.

Krieghbaum, H. (Ed.). (1963). Special section on science news. *Journalism Quarterly*, *40*, 291–338.

LaFollette, M. C. (1990, July 9). Scientists and the media: In search of a healthier symbiosis. *The Scientist*, 13, 15.

Latour, B., & Woolgar, S. (1979). *Laboratory life.* Beverly Hills, CA: Sage.

Lewenstein, B. V. (1988). Davis, Watson. In J. A. Garraty & M. C. Carnes (Eds.). *Dictionary of American biography* (pp. 119–120). New York: Charles Scribner's Sons.

Lewenstein, B. V. (1992). Cold fusion and hot history. *Osiris, 7,* 135–163.

Lewenstein, B. V. (1995a). Do public electronic bulletin boards help create scientific knowledge? The cold fusion case. *Science, Technology & Human Values, 20*(2), 123–49.

Lewenstein, B. V. (1995b). From fax to facts: Communication in the cold fusion saga. *Social Studies of Science, 25* (3), 403–436.

Lewenstein, B. V. (1997). Life on Mars—and in science. *Mercury*, *26*(1), 24–26.

Lewenstein B. V., & Baur, W. (1991). A cold fusion chronology. *Journal of Radioanalytical and Nuclear Chemistry*, *152,* 279–298.

Lewenstein, B. V., & Heinz, L. (1992, December). *Changing culture of research: Processes of knowledge transfer.* Contractor report to U.S. Congress, Office of Technology Assessment.

Mallove, E. (1991). *Fire from ice: Searching for the truth behind the cold fusion furor.* New York: John Wiley.

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations* (Norman W. Storer, Ed.). Chicago: University of Chicago Press.

Mulkay, M. (1979). *Science and sociology of knowledge.* London: George Allen & Unwin.

National Academy of Sciences. (1959). *Proceedings of the International Conference on Scientific Information.* Washington, DC: Author.

Nelkin, D. (1985). *Selling science: How the press covers science and technology* (Rev. ed.). New York: W. H. Freeman.

Nelson, C. E., & Pollack, D. K. (Eds.). (1970). *Communication among scientists and engineers.* Lexington, MA: Heath Lexington Books.

Nowotny, H., & Felt, U. (1997). *After the breakthrough: The emergence of high-temperature superconductivity as a research field.* New York: Cambridge University Press.

Petrasso, R. D., Chen, X., Wenzel, K. W., Parker, R. R., Li, C. K., Fiore, C. (1989, 18 May). Problems with the gamma-ray spectrum in the Fleischmann et al. experiments. *Nature*, *339*, 183–185.

Phillips, D. P., Kanter, E. J., Bednarczyk, B.,& Tastad, P. L. (1991). Importance of the lay press in the transmission of medical knowledge to the scientific community. *New England Journal of Medicine*, *325,* 1180–1183.

Pool, R. (1989, June 2). Cold fusion: End of act I. *Science*, *245,* 1039–1040.

Ravetz, J. R. (1971). *Scientific knowledge and its social problems.* Oxford: Clarendon Press.

Royal Society. (1948). *The Royal Society Scientific Information Conference, 21 June–2 July 1948: Report and papers submitted.* London: Author.

Salamon, M. H., Wrenn, M. E., Bergeson, H. E., Crawford, K. C, Delaney, W. H., Henderson, C. L., Li, Y. Q., Rusho, J. A., Sandquist, G. M., Seltzer, S. M. (1990, March 29). Limits on the emission of neutrons, gamma-rays, electrons and protons from Pons/Fleischmann electrolytic cells. *Nature*, *344*, 401–405.

Schechter, B. (1989). *The path of no resistance*. New York: Simon and Schuster.

Shannon, C., & Weaver, W. (1949). *The mathematical theory of communication.* Urbana: University of Illinois Press.

Shinn, T., & Whitley R. (Eds.). (1985). *Expository science: Forms and functions of popularisation.* Dordrecht/Boston/Lancaster: D. Reidel.

Smith, D. (1989). Quest for fusion. *[Caltech] Engineering and Science*, *Summer*, 3–15.

Taubes, G. (1990, June 15). Cold fusion conundrum at Texas A&M. *Science*, *248*, 1299.

Taubes, G. (1993). *Bad science: The short life and very hard times of cold fusion.* New York: Random House.

University Microfilms, Inc. *Newspaper Abstracts Online.* Ann Arbor, MI: author.

Will, F., et al. (Eds.). (1990). *The first annual conference on cold fusion: Conference proceedings.* Salt Lake City, UT: National Cold Fusion Institute.

Ziman, J. M. (1968). *Public knowledge: An essay concerning the social dimension of science.* Cambridge: Cambridge University Press.

Shaping Biomedicine as an Information Science

Timothy Lenoir

A New Biology for the Information Age

Sometime in the mid-1960s biology became an information science. While François Jacob and Jacques Monod's work on the genetic code is usually credited with propelling biology into the Information Age, in this essay I explore the transformation of biology by what have become essential tools to the practicing biochemist and molecular biologist: namely, the contributions of information technology. About the same time as Jacob and Monod's work, developments in computer architectures and algorithms for generating models of chemical structures and simulations of chemical interactions were created that allowed computational experiments to interact with and draw together theory and laboratory experiment in completely novel ways. The new computational science linked with visualization has had a major impact in the fields of biochemistry, molecular dynamics, and molecular pharmacology (Friedhoff & Benzon, 1989; Panel on Information Technology and the Conduct of Research, 1989; McCormick, DeFanti, & Brown, 1987, p. A-1; Hall, 1995). By "computational science" I mean the use of computers in science disciplines like these as distinct from computer science (McCormick, DeFanti, & Brown, p. 11). The sciences of visualization are defined by McCormick, DeFanti, and Brown as follows:

> Images and signals may be captured from cameras or sensors, transformed by image processing, and presented pictorially on hard or soft copy output. Abstractions of these visual representations can be transformed by computer vision to create symbolic representations in the form of symbols and structures. Using computer graphics, symbols or structures can be synthesized into visual representations. (P. A-1)

Computational approaches have substantially transformed and extended the domain of theorizing in these areas in ways unavailable to older, non–computer-based forms of theorizing.

But other information technologies have also proved crucial to bringing about this change. In the 1970s through the 1990s, armed with such new tools of molecular biology as cloning, restriction enzymes, protein sequencing, and gene product amplification, biologists were awash in a sea of new data. They deposited this data in large and growing electronic databases of genetic maps, atomic coordinates for chemical and protein structures, and protein sequences. These developments in technique and instrumentation launched biology onto the path of becoming a data-bound science, a "science" in which *all* the data of a domain—such as a genome—are available before the laws of the domain are understood. Biologists have coped with this data explosion by turning to information science: applying artificial intelligence and expert systems and developing search tools to identify structures and patterns in their data.

The aim of this paper is to explore early developments in the introduction of computer modeling tools from artificial intelligence (AI) and expert systems into biochemistry in the 1960s and 1970s, and the introduction of informatics techniques for searching databases and extracting biological function and structure in the emerging field of genomics during the 1980s and 1990s. I have two purposes in this line of inquiry. First I want to suggest that by introducing tools of information science biologists have sought to make biology a unified theoretical science with predictive powers analogous to other theoretical disciplines. But I want also to suggest that along with this highly heterogeneous and hybrid form of computer-based experimentation and theorizing has come a different conception of theorizing itself: one based on models of information-processing and best captured by the phrase "knowledge engineering" developed within the AI community. My second concern is to contribute to recent discussions on the transformation of biology into an information science. Lily Kay, Evelyn Fox Keller, Donna Haraway, and Richard Doyle have explored the role of metaphor, disciplinary politics, economics, and culture in shaping the context in which the language of "DNA code," "genetic information," "text," and "transcription" have been inserted into biological discourse, often in the face of resistance from some of the principal actors themselves (Doyle, 1997). I am more interested than these authors in software and the computational regimes that it enables. Elaborating on the theme of "tools to theory," recently espoused in science and technology studies, I am interested in exploring the role of the computational medium itself in shaping biology as an information science. But a further crucial stimulation to the takeoff of bioinformatics, of course, was provided by hardware and networking developments underwritten by the NIH and NSF (Hughes, 1999).

Computers and Biochemistry: Molecular Modeling

The National Institutes of Health have been active at every stage in making biology an information science. NIH support was crucial to the explosive take-off of computational chemistry and the promotion of computer-based visualization technologies in the mid-1960s. The agency sponsored a conference at UCLA in 1966 on "Image Processing in Biological Science." The NIH's Bruce Waxman, co-chair of the meeting, set out the NIH agenda for computer visualization by sharply criticizing the notion of mere "image processing" as the direction that should be pursued in computer-enhanced vision research. The goal of computer-assisted "vision," he asserted, was not to replicate relatively low-order motor and perceptual capabilities even at rapid speeds. "I have wondered whether the notion of image processing is itself restrictive; it may connote the reduction of and analysis of 'natural' observations but exclude from consideration two- or three-dimensional data which are abstractions of phenomena rather than the phenomena themselves" (Ramsey, 1968, pp. xiii–xiv). Waxman suggested "pattern recognition" as the subject that they should really pursue—and in particular where the object was what he termed "non-natural." In general, Waxman asserted, by its capacity to quantize massive data sets automatically, the computer, linked with pattern-recognition methods of imaging the non-natural, would permit the development of stochastically based biological theory.

Waxman's comments point to one of the important and explicit goals of the NIH and other funding agencies: to mathematize biology. That biology should follow in the footsteps of physics had been the centerpiece of a reductionist program since at least the middle of the nineteenth century. But the development of molecular biology in the 1950s and 1960s encouraged the notion that a fully quantitative theoretical biology was on the horizon. The computer was to be the motor for this change. Analogies were drawn between highly mathematized and experimentally based Big Physics and the anticipated "Big Biology." As Lee B. Lusted, the chairman of the Advisory Committee to the National Research Council on Electronic Computers in Biology and Medicine argued, because of the high cost of computer facilities for conducting biological research, computer facilities would be to biology what SLAC (the Stanford Linear Accelerator) and the Brookhaven National Laboratory were to physics (Ledley, 1965, pp. ix–x). Robert Ledley, then affiliated with the Division of Medical Sciences, National Research Council, and author of the volume expressed the committee's interest in fostering computing and insisted that biology was on the threshold of a new era. New emphasis on quantitative work and experiment was changing the conception of the biologist: the view of the biologist as an individual scientist, personally carrying through each step of his investigation and his data-reduction processes, was rapidly broadened to include the biologist as a part of an intricate organizational chart that partitions scientific, technical, and administrative responsibilities. In the new organization of biological work, modeled on the physicists' work at large national labs, the talents and knowledge of the biologist "must be augmented by those of the engineer, the mathematical analyst, and the professional computer programmer" (Ledley, 1965, p. xi).

At the UCLA meeting Bruce Waxman held up as a model the work on three-dimensional representations of protein molecules carried out by Cyrus Levinthal. Levinthal worked with the facilities of MIT's Project on Mathematics and Computation (MAC), one of the first centers in the development of graphics. Levinthal's project was an experiment in computer time-sharing linking biologists, engineers, and mathematicians in the construction of Big Biology. Levinthal's work at MIT illustrates the role of computer visualization as a condition for theory development in molecular biology and biochemistry.

Since the work of Linus Pauling and Robert Corey on the α-helical structure of most protein molecules in 1953, models have played a substantial role in biochemistry. Watson and Crick's construction of the double helix model for DNA depended crucially upon the construction of a physical model. Subsequently, work in the field of protein biology has demonstrated that the functional properties of a molecule depend not only on the interlinkage of its chemical constituents but also on the way in which the molecule is configured and folded in three dimensions. Much of biochemistry has focused on understanding the relationship between biological function and molecular conformational structure.

A milestone in the making of physical models (in three dimensions) of molecules was John Kendrew's construction of myoglobin. The power of models in investigations of biomolecular structure was evident from work such as this, but such tools had limitations as well. Kendrew's model, for instance, was the first successful attempt to build a physical model into a Fourier map of a molecule's electron densities derived from X-ray crystallographic sources. As a code for electron density, clips of different colors were put at the proper vertical positions on a forest of steel rods. A brass wire model of the alpha helices and beta sheets that make up the molecule was then built in among the rods. Mechanical interference made it difficult to adjust the structure, and the model was hard to see because of the large number of supporting rods. The model incorporated both too little and too much: too little, in that the basic shape of the molecule was not represented; too much, in that the forest of rods made it difficult to see the three-dimensional folding of the molecule (even though bond connectivity was represented). Perhaps the greatest drawback was the model's size: It filled a large room. The answer to these problems was computer representation. For an early stereogram of myoglobin constructed by computer on the basis of Kendrew's data, see Watson (1969). It was obvious that such three-dimensional representations would only become really useful when it was possible to manipulate them at will. Proponents of computer graphics argued that this flexibility is exactly what computer representations of molecular structure would allow. Cyrus Levinthal first illustrated these methods in 1965.

Levinthal reasoned that since protein chains are formed by linking molecules of a single class, amino acids, it should be relatively easy to specify the linkage process in a form mathematically suitable for a digital computer (Levinthal, 1966). Initially the computer model considers the molecule as a set of rigid groups of constant geometry linked by single bonds around which rotation is possible. Program input consists of a set of coordinates consistent with the molecular stereochemistry as given in data from X-ray crystallographic studies. Several constraints delimit stable configurations among numerous possibilities resulting from combinations of linkages among the twenty different amino acids. These include bond angles, bond lengths, van der Waals radii for each species of atom, and the planar configuration of the peptide bond.

Molecular biologists, particularly the biophysicists among them, were motivated to build a unified theory, and the process of writing a computer program that could simulate protein structure would assist in this goal by providing a framework of mental and physical discipline from which would emerge a fully mathematized theoretical biology. In such non-mathematized disciplines as biology, the language of the computer program would serve as the language of science (Oettinger, 1966, p. 161). But there was a hitch: In an ideal world dominated by a powerful central theory, one would like, for example, to use the inputs of xyz coordinates of the atoms, types of bond, and so forth, to calculate the pairwise interaction of atoms in the amino acid chain, predict the conformation of the protein molecule, and check this against its corresponding X-ray crystallographic image. As described, however, the parameters used as input in the computer program do not provide much limitation on the number of molecular conformations. Other sorts of input are needed to filter among the myriad possible structures. Perhaps the most important of these is energy minimization. In explaining how the thousands of atoms in a large protein molecule interact with one another to produce a stable conformation, one hypothesizes that, like water running downhill, the molecular string will fold to reach a lowest energy level. To carry out this sort of minimization would entail calculating the interactions of all pairs of active structures in the chain, minimizing the energy corresponding to these interactions over all possible configurations, and then displaying the resulting molecular picture. Unfortunately, this objective could not be achieved, as Levinthal noted, because a formula describing such interactions could not, given the state of molecular biological theory in 1965, even be stated, let alone be manipulated with a finite amount of labor. In Levinthal's words:

> The principal problem, therefore, is precisely how to provide correct values for the variable angles. . . . I should emphasize the magnitude of the problem that remains even after one has gone as far as possible in using chemical constraints to reduce the number of variables from

> several thousand to a few hundred. . . . I therefore decided to develop programs that would make use of a man-computer combination to do a kind of model-building that neither a man nor a computer could accomplish alone. This approach implies that one must be able to obtain information from the computer and introduce changes in the way the program is running in a span of time that is appropriate to human operation. This in turn suggests that the output of the computer must be presented not in numbers but in visual form. (Levinthal, 1966, pp. 48–49)

In Levinthal's view, visualization generated in real-time interaction between human and machine can assist theory construction. The computer becomes in effect both a microscope for examining molecules as well as a laboratory for quantitative experiment. Levinthal's program, CHEMGRAF, could be programmed with sufficient structural information as input from physical and chemical theory to produce a trial molecular configuration as graphical output. A subsystem called SOLVE then packed the trial structure by determining the local minimum energy configuration due to non-bonded interactive forces. A subroutine of this program, called ENERGY, calculated the torque vector caused by the atomic pair interactions on rotatable bond angles. An additional procedure for determining the conformation of the model structure was "cubing." This procedure searched for nearest neighbors of an atom in the center of a $3 \times 3 \times 3$ cube and reported whether any atoms were in the twenty-six adjacent cubes. The program checked for atom pairs in the same or adjacent cubes and for atoms within a specified distance. It maintained a list of pairs that were, for instance, in contact violation, while another routine calculated energy contribution of the pair to the molecule. The cubing program rejected as early as possible all those atom pairs where the interatomic distance was too great to be of more than negligible contribution, and it enabled more efficient use of computer time.

Levinthal emphasized that interactivity was a crucial component of CHEMGRAF. Built into his system was the requirement of observing the result of the calculations interactively so that one could halt the minimization process at any step, either to terminate it completely or to alter the conformation and then resume it (Katz & Levinthal, 1972). Levinthal noted that often, as the analytical procedures were grinding on, a molecule would be trapped in an unfavorable conformation or in a local minimum and the problem would be obscure until the conformation could be viewed three-dimensionally. CHEMGRAF enabled the investigator to assist in generating the local minimization of energy for a subsection of the molecule through three different types of user-guided empirical manipulation of structure: "close," "glide," and "revolve." These manipulations in effect introduced external "pseudo-energy" terms into the computation that pulled the structure in various ways (Levinthal, Barry, Ward, & Zwick, 1968). Atoms could be rotated out of the way by direct command and a new starting conformation chosen from which to continue the minimization procedure. By pulling individual atoms to specific locations indicated by experimental data from X-ray diffraction studies, a fit between X-ray crystallographic data and the computer model of a specific protein, such as myoglobin, could ultimately be achieved. With the model in hand of the target molecule, such as myoglobin, one could then proceed to investigate the various energy terms involved in holding the protein molecule together. Thus, the goal of this interactive effort involving human and machine was eventually to generate a theoretical formulation for the lowest energy state of a protein molecule, to predict its structure, and to have that prediction confirmed by X-ray crystallographic images (Hall, 1995).

The enormous number of redundant trial calculations involved in Levinthal's work hints at the desirability of combining an expert system with a visualization system. E. J. Corey and W. Todd Wipke, working nearby at Harvard, took this next step. (Space limitations prevent me from discussing their work here.) In developing their work, Wipke and Corey drew upon a prototype expert system at Stanford called DENDRAL, the result of a collaboration at Stanford among computer scientist Edward Feigenbaum, biologist Joshua Lederberg, and organic chemist Carl Djerassi, working on another of the NIH initiatives to bring computers directly into the laboratory. The Stanford project, called DENDRAL, was an early effort in the field of what Feigenbaum and his mentors Herbert Simon and Marvin Minsky termed "knowledge engineering." In effect, it attempted to put the human inside the machine.

DENDRAL: The AI Approach at Stanford

DENDRAL aimed at emulating an organic chemist operating in the harsh environment of Mars (Lederberg, n.d.; Lederberg, Sutherland, Buchanan, & Feigenbaum, 1969). The ultimate goal was to create an automated laboratory as part of the Viking mission planned to land a mobile instrument pod on Mars in 1975. Given the mass spectrum of an unknown compound, the specific goal was to determine the structure of the compound. To accomplish this, DENDRAL would analyze the data, generate a list of plausible candidate structures, predict the mass spectra of those structures from the theory of mass spectrometry, and select as a hypothesis the structure whose spectrum most closely matched the data.

A key part of this program was the representation of chemical structure in terms of topological graph theory. Chemical graphs were the visual "language" to augment the theoretical and practical knowledge of the chemist with the calculating power of the computer. This part of the effort was contributed by Lederberg, the winner of the 1958 Nobel Prize in medicine or physiology, for his work on genetic exchange in bacteria, who had been interested in the introduction of information concepts into biology for most of his professional life. Self-described as a man with a Leibnizian dream of a universal calculus for the alphabet of human thought, Lederberg's interest in mass spectrometry and topological mapping of molecules was in part driven by the dream of mathematizing biology, starting with organic chemistry. The structures of organic molecules are bewilderingly complex, and the "theory" of organic chemistry does not have an elegant axiomatic structure analogous, say, to Newtonian mechanics, even though it is sprinkled with lots of theory derived from quantum mechanics and thermodynamics. Lederberg felt that a first step toward such a quantitative, predictive theory would be a rational systematization of organic chemistry. Trampling upon a purist's notion of theory, Lederberg thought that computers were the royal road to mathematization in chemistry:

> Could not the computer be of great assistance in the elaboration of novel and valid theories? I can dream of machines that would not only execute experiments in physical and chemical biology but also help design them, subject to the managerial control and ultimate wisdom of their human programmer. (Lederberg, 1969)

Mass spectrometry, the area upon which Feigenbaum and Lederberg concentrated with Carl Djerassi, was a particularly appropriate challenge. It differed in at least one crucial aspect from the molecular modeling of proteins I have considered above. Whereas in those areas a well-understood theory, such as the quantum mechanical theory of the atomic bond, was the basis for developing the computer program to examine effects in large calculations, there was no theory of mass spectrometry that could be transferred to the program from a textbook (Lederberg, Sutherland, Buchanan, & Feigenbaum, 1969). The field has bits of theory to draw upon, but it has developed mainly by following rules of thumb, which are united in the form of the chemist-expert. The field thrives on tacit knowledge. The following excerpt from a memo by Feigenbaum written after his first meetings with Lederberg on the DENDRAL project provides a vivid sense of the objective and the problems faced:

> The main assumption we are operating under is that the required information is buried in chemists' brains if only we can extract it. Therefore, the initiative for the interaction must come from the system not the chemist, while allowing the chemist the flexibility to supply additional information and to modify the question sequence or content of the system. . . . What we want to design then is a question asking system [that] will gather rules about the feasibility of the chemical molecules and their subgraphs being displayed. ("Second Cut," n.d.)

In short, Feigenbaum sought to emulate a gifted chemist with the computer. That chemist was Carl Djerassi, nicknamed "El Supremo" by his graduate and postdoctoral students. Djerassi's astonishing achievements as a mass spectrometrist relied on his abilities to feel his way through the process without the aid of a complete theory, relying rather on experience, tacit knowledge, hunches, and rules of thumb. In interviews Feigenbaum elicited this kind of information from Djerassi, in a process that heightened awareness of the structure of the field for both participants. The process of involving a computer in chemical research in this way organized a variety of kinds of information, which constituted a crucial step toward theory.

A Paradigm Shift in Biology

Thus far I have been considering efforts to predict structure from physical principles as the first path through which computer science and computer-based information technology began to reshape biology. The Holy Grail of biology has always been the construction of a mathematized theoretical biology, and for most molecular biologists the journey there has been directed by the notion that the information for the three-dimensional folding and structure of proteins is uniquely contained in the linear sequence of their amino acids (Anfinsen, 1973). As we have seen, the molecular dynamics approach assumed that if all the forces between atoms in a molecule, including bond energies and electrostatic attraction and repulsion, are known, then it is possible to calculate the three-dimensional arrangement of atoms that requires the least energy. Christian B. Anfinsen (1973) discussed the work for which he was awarded the Nobel Prize in chemistry in 1972:

> This hypothesis (the "thermodynamic hypothesis") states that the three-dimensional structure of a native protein in its normal physiological milieu . . . is the one in which the Gibbs free energy of the whole system is lowest; that is, that the totality of interatomic interactions and hence by the amino acid sequence, in a given environment. (P. 223)

Because this method requires intensive computer calculations, shortcuts have been developed that combine computer-intensive molecular dynamics computations, artificial intelligence, and interactive computer graphics in deriving protein structure directly from chemical structure.

While theoretically elegant, the determination of protein structure from chemical and dynamical principles has been hobbled with difficulties. In the abstract, analysis of physical data generated from protein crystals, such as X-ray and nuclear magnetic resonance data, should offer rigorous ways to connect primary amino acid sequences to three-dimensional structure. But the problems of acquiring good crystals and the difficulty of getting NMR data of sufficient resolution are impediments to this approach. Moreover, while quantum mechanics provides a solution to the protein-folding problem in theory, the computational task of predicting structure from first principles for large protein molecules containing many thousands of atoms has proved impractical. Furthermore, unless it is possible to grow large, well-ordered crystals of a given protein, X-ray structure determination is not an option. The development of methods of structure determination by high-resolution two-dimensional NMR has alleviated this situation somewhat, but this technique is also costly and time-consuming, requiring large amounts of protein of high solubility, and is severely limited by protein size. These difficulties have contributed to the slow rate of progress in registering atomic coordinates of macromolecules.

An indicator of the difficulty of pursuing this approach alone is suggested by the relatively slow growth of databanks of atomic coordinates for proteins. The Protein Data Bank (PDB) was established in 1971 as a computer-based archival resource for macromolecular structures. The purpose of the PDB was to collect, standardize, and distribute atomic coordinates and other data from crystallographic studies. In 1977 the PDB listed atomic coordinates for forty-seven macromolecules (Bernstein et al., 1977). In 1987 that number began to increase rapidly at a rate of about 10 percent per year because of the development of area detectors and widespread use of synchrotron radiation; by April 1990 atomic coordinate entries existed for 535 macromolecules. Commenting on the state of the art in 1990, Holbrook and colleagues (1993) noted that crystal determination could require one or more man-years. Currently (1999), the PDB's Biological Macromolecule Crystallization Database (BMCD) contains entries for 2,526 biological macromolecules for which diffraction quality crystals have been obtained. These include proteins, protein:protein complexes, nucleic acid, nucleic acid:nucleic acid complexes, protein:nucleic acid complexes, and viruses.[1]

While structure determination was moving at a snail's pace, beginning in the 1970s, another stream of work contributed to the transformation of biology into an information science. The development of restriction enzymes, recombinant DNA techniques, gene cloning techniques, and polymerase chain reactions (PCRs) resulted in a flood of data on DNA, RNA, and protein sequences. Indeed more than 140,000 genes were cloned and sequenced in the twenty years from 1974 to 1994, of which more than 20 percent were human genes (Brutlag, 1994, p. 159). By the early 1990s, well before the beginning of the Human Genome Initiative, the NIH GenBank database (release 70) contained more than 74,000 sequences, while the Swiss Protein database (Swiss-Prot) included nearly 23,000 sequences. Protein databases were doubling in size every twelve months, and some were predicting that by the year 2000 ten million base pairs a day would be sequenced as a result of the technological impact of the Human Genome Initiative. Such an explosion of data encouraged the development of a second approach to determining the function and structure of protein sequences: namely, prediction from sequence data alone. This "bioinformatics" approach identifies the function and structure of unknown proteins by applying search algorithms to existing protein libraries in order to determine sequence similarity, percentages of matching residues, and the statistical significance of each database sequence.

A key project illustrating the ways in which com-

[1] Biological Macromolecule Crystallization Database and the NASA Archive for Protein Crystal Growth Data (version 2.00) are located on the Web at http://www.bmcd.nist.gov:8080/bmcd/bmcd.html.

puter science and molecular biology began to merge in the formation of bioinformatics was the MOLGEN project at Stanford and events related to the formation and subsequent development of BIONET. MOLGEN was a continuation of the projects in artificial intelligence and knowledge engineering begun at Stanford with DENDRAL. MOLGEN was started in 1975 as a project in the Heuristic Programming Project with Edward Feigenbaum as principal investigator directing the thesis projects of Mark Stefik and Peter Friedland (Feigenbaum & Martin, 1977). The aim of MOLGEN was to model the experimental design activity of scientists in molecular genetics (Friedland, 1979). Before an experimentalist sets out to achieve some goal, he produces a working outline of the experiment, guiding each step of the process. The central idea of MOLGEN was based on the observation that scientists rarely plan from scratch in designing a new experiment. Instead, they find a skeletal plan, an overall design that has worked for a related or more abstract problem, and then adapt it to the particular experimental context. Like DENDRAL, this approach is heavily dependent upon large amounts of domain-specific knowledge in the field of molecular biology and even more upon good heuristics for choosing among alternative implementations.

MOLGEN's designers chose molecular biology as appropriate for the application of artificial intelligence because the techniques and instrumentation generated in the 1970s seemed ripe for automation. The advent of rapid DNA cloning and sequencing methods had had an explosive effect on the amount of data that could be most readily represented and analyzed by a computer. Moreover, it appeared that very soon progress in analyzing information in DNA sequences would be limited by the lack of an appropriate combination of search and statistical tools. MOLGEN was intended to apply rules to detect profitable directions for analysis and to reject unpromising ones (Feigenbaum et al., 1980).

Peter Friedland was responsible for constructing the knowledge-base component of MOLGEN. Though not himself a molecular biologist, he made a major contribution to the field by assembling the rules and techniques of molecular biology into an interactive, computerized system of analytical programs. Friedland worked with Stanford molecular biologists Douglas Brutlag, Laurence Kedes, John Sninsky, and Rosalind Grymes, who provided expert knowledge on enzymatic methods, nucleic acid structures, detection methods, and pointers to key references in all areas of molecular biology. Along with providing an effective encyclopedia of information about technique selection in planning a laboratory experiment, the knowledge base contained a number of tools for automated sequence analysis. Brutlag, Kedes, Sninsky, and Grymes were interested in having a battery of automated tools for sequence analysis, and they contracted with Friedland and Stefik—both gifted computer program designers—to build these tools in exchange for contributing their expert knowledge to the project (Douglas Brutlag, personal communication; Peter Friedland, personal communication). (In 1987, after his work on MOLGEN and at IntelliGenetics [discussed below], Friedland went on to become chief scientist at the NASA-Ames Laboratory for Artificial Intelligence.)

This collaboration of computer scientists and molecular biologists helped move biology along the road to becoming an information science. Among the programs Friedland and Stefik created for MOLGEN was SEQ, an interactive self-documenting program for nucleic acid sequence analysis, which had thirteen different procedures with over twenty-five different subprocedures, many of which could be invoked simultaneously to provide various analytical methods for any sequence of interest. SEQ brought together in a single program methods for primary sequence analysis described in the literature by L. J. Korn and colleagues, R. Staden, and numerous others (Korn, Queen, & Wegman, 1977); Staden, 1977; Staden, 1978; Staden, 1979). SEQ also performed homology searches on DNA sequences and specified the degree of homology, and conducted dyad symmetry (inverted repeats) searches (Friedland, Brutlag, Clayton, & Kedes, 1982). Another feature of SEQ was its ability to prepare restriction maps with the names and locations of the restriction sites marked on the nucleotide sequence. In addition it had a facility for calculating the length of DNA fragments from restriction digests of any known sequence. Another program in the MOLGEN suite was GA1 (later called MAP). Constructed by Stefik, GA1 was an artificial intelligence program that allowed the generation of restriction enzyme maps of DNA structures from segmentation data (Stefik, 1977). It would construct and evaluate all logical alternative models that fit the data and rank them in relative order of fit. A further program in MOLGEN was SAFE, which aided in enzyme selection for gene excision. SAFE took amino acid sequence data and predicted the restriction enzymes guaranteed not to cut within the gene itself.

In its first phase of development (1977–1980) MOLGEN consisted of the programs described above and a knowledge base containing information on about three hundred laboratory methods and thirty strategies for using them. It also contained the best currently available data on about forty common phages, plasmids, genes, and other known nucleic acid structures. The second phase of development beginning in 1980 scaled up the analytical tools and the knowledge base. Perhaps the most significant aspect of the second phase was making MOLGEN available to the scientific community at large on the Stanford University Medical Experimental national computer resource, SUMEX-AIM. SUMEX-AIM, supported by the Biotechnology Resources Program at NIH since 1974, had been home to DENDRAL and several other programs. The new experimental resource on SUMEX, comprising the MOLGEN programs and access to all major genetic databases, was called GENET. In February 1980 GENET was made available to a carefully limited community of users (Rindfleisch, Friedland, & Clayton, 1981).

MOLGEN and GENET were immediate successes with the molecular biology community. In their first few months of operation in 1980 more than two hundred labs (with several users in each of those labs) accessed the system. By 1 November 1982 more than three hundred labs on the system around the clock accessed the system from a hundred institutions (Douglas Brutlag, personal communication; NIH Special Study Section, 1983; Lewin, 1984). Traffic on the site was so heavy that restrictions had to be implemented and plans for expansion considered. In addition to the academic users a number of biotech firms, such as Monsanto, Genentech, Cetus, and Chiron, used the system heavily. Feigenbaum, principal investigator in charge of the SUMEX resource, and Thomas Rindfleisch, facility manager, decided to exclude commercial users in order to ensure that the academic community had unrestricted access to the SUMEX computer and to answer the NIH's concern that commercial users gain unfair access to the resource (Maxam to GENET community, 1982).

To provide commercial users with their own unrestricted access to the GENET and MOLGEN programs, Brutlag, Feigenbaum, Friedland, and Kedes formed a company, IntelliGenetics, which would offer the suite of MOLGEN software for sale or rental to the emerging biotechnology industry. With 125 research labs doing recombinant DNA research in the United States alone and a number of new genetic engineering firms starting up, opportunities looked outstanding. No one was currently supplying software in this rapidly growing genetic engineering marketplace. With their exclusive licensing arrangement with Stanford for the MOLGEN software, IntelliGenetics was poised to lead a huge growth area. The business plan expressed well the excellent position of the company:

> A major key to the success of IntelliGenetics will be the fact that the recombinant DNA research revolution is so recent. While every potential customer is well capitalized, few have the manpower they say they need; this year several firms are hiring 50 molecular geneticist Ph.D.s, and one company speaks of 1000 within five years. These firms require computerized assistance—for the storage and analysis of very large amounts of DNA sequence information which is growing at an exponential rate—and will continue to do so for the foreseeable future (10 years). Access to this information and the ability to perform rapid and efficient pattern recognition among these sequences is currently being demanded by most of the firms involved in recombinant DNA research.
>
> The programs offered by IntelliGenetics will enable the researchers to perform tasks that are: 1) virtually impossible to perform with hand calculations, and 2) extremely time-consuming and prone to human error. *In other words, IntelliGenetics offers researcher productivity improvement to an industry with expanding demand for more researchers which is experiencing a severe supply shortage* [emphasis in original]. ("Business plan for IntelliGenetics," 1981; Friedland to Reimers, 1984)[2]

The resource that IntelliGenetics eventually offered to commercial users was BIONET. Like GENET, BIONET combined all databases of DNA sequences with programs to aid in their analysis in one computer site.

Prior to the startup of BIONET and contemporaneous with GENET, other resources for DNA sequences were developed. Several researchers were making their databases available. Under the auspices of the National Biomedical Research Foundation, Margaret Dayhoff had created a database of DNA sequences and some software for sequence analysis that was marketed commercially. Walter Goad, a physicist at Los Alamos National Laboratory, collected DNA sequences from the published literature and made them freely available to researchers. But by the late 1970s the number of bases sequenced was already approaching three million and expected to

[2] Details of the software licensing arrangement and the revenues generated are discussed in a letter to Niels Reimers, at the Stanford Office of Technology Licensing, on the occasion of renegotiating the terms.

double soon. Some form of easy communication between labs for effective data handling was considered a major priority in the biological community. While experiments were going on with GENET, a number of nationally prominent molecular biologists had been pressing to start an NIH-sponsored central repository for DNA sequences. In 1979, at Rockefeller University, Joshua Lederberg organized an early meeting with such an agenda. The proposed NIH initiative was originally supposed to be coordinated with a similar effort at the European Molecular Biology Laboratory (EMBL) in Heidelberg, but the Europeans became dissatisfied with the lack of progress on the American end and decided to go ahead with their own databank. EMBL announced the availability of its Nucleotide Sequence Data Library in April 1982, several months before the American project was funded. Finally, in August 1982, the NIH awarded a contract for $3 million over five years to the Boston-based firm of Bolt, Berenek, and Newman (BB&N) to set up the national database known as GenBank in collaboration with Los Alamos National Laboratory. IntelliGenetics submitted an unsuccessful bid for that contract.

The discussions leading up to GenBank included consideration of funding a more ambitious databank, known as "Project 2," which was to provide a national center for the computer analysis of DNA sequences. Budget cuts forced the NIH to abandon that scheme (Lewin, 1984). However, officials there returned to it the following year, thanks to the persistence of IntelliGenetics representatives. Although GenBank launched a formal national DNA sequence collection effort, the need for computational facilities voiced by molecular biologists was still left unanswered. In September 1983, after a review process that took over a year, the NIH division of research resources awarded IntelliGenetics a $5.6 million five-year contract to establish BIONET (Lewin, 1984). The contract, the largest award of its kind by the NIH to a for-profit organization (p. 1380), started on 1 March 1984 and ended on 27 February 1989.

BIONET first became available to the research community in November 1984. The fee for use was $400 per year per laboratory and remained at that level throughout its first five years. BIONET's use grew impressively. Initially the IntelliGenetics team set the target for user subscriptions at 250 labs. However, in March 1985, the annual report for the first year's activities of BIONET listed 350 labs with nearly 1,132 users. By August 1985 that number had increased dramatically to 450 labs and 1,500 users (Minutes of the meeting, 1985). In April 1986, for example, BIONET had 464 laboratories comprising 1,589 users. By October 1986 the numbers were 495 labs and 1,716 users (BIONET users status, 1986). By 1989, 900 laboratories in the United States, Canada, Europe, and Japan (comprising about 2,800 researchers) subscribed to BIONET, and 20 to 40 new laboratories joined each month (Huberman, 1989).

BIONET was intended to establish a national computer resource for molecular biology satisfying three goals, which it fulfilled to varying degrees. A first goal was to provide a way for academic biologists to obtain access to computational tools to facilitate research relating to nucleic acids and possibly proteins. In addition to giving researchers ready access to national databases on DNA and protein sequences, BIONET would provide a library of sophisticated software for sequence searching, matching, and manipulation. A second goal was to provide a mechanism to facilitate research into improving such tools. The BIONET contract provided research and development support of further software, both in-house research by IntelliGenetics scientists and through collaborative ventures with outside researchers. A third goal of BIONET was to enhance scientific productivity through electronic communications.

The stimulation of collaborative work through electronic communication was perhaps the most impressive achievement of BIONET. BIONET was much more than the Stanford GENET plus the MOLGEN–IntelliGenetics suite of software. Whereas GENET with its pair of ports could accommodate only two users at any one time, BIONET had twenty-two ports providing an estimated annual thirty thousand connect hours (Friedland, 1984; Smith, Brutlag, Friedland, & Kedes, 1986). All subscribers to BIONET were provided with e-mail accounts. For most molecular biologists this was something entirely new, since most university labs were just beginning to be connected with regular e-mail service. At least twenty different bulletin boards on numerous topics were supported by BIONET. In an effort to change the culture of molecular biologists by accustoming them to the use of electronic communications and more collaborative work, BIONET users were required to join one of the bulletin board groups.

BIONET subscribers had access to the latest versions of the most important databases for molecular biology. Large databases available at BIONET were GenBank, the National Institutes of Health DNA sequence library; EMBL, the European Molecular Biology

Laboratory nucleotide sequence library; NBRF-PIR, the National Biomedical Research Foundation's protein sequence database, which is part of the Protein Identification Resource [PIRI] supported by NIH's Division of Research Resources; SWISS-PROT, a protein sequence database founded by Amos Bairoch of the University of Geneva and subsequently managed and distributed by the European Molecular Biology Laboratory; VectorBank, IntelliGenetics' database of cloning vector restriction maps and sequences; Restriction Enzyme Library, a complete list of restriction enzymes and cutting sites provided by Richard Roberts at Cold Spring Harbor; and Keybank, IntelliGenetics' collection of predefined patterns or "keys" for database searching. Several smaller databases were also available, including a directory of molecular biology databases, a collection of literature references to sequence analysis papers, and a complete set of detailed protocols for use in a molecular biological laboratory (especially for *Escherichia coli* and yeast work) (IntelliGenetics, 1987, p. 23).

Perhaps the most important contribution made by BIONET to establishing molecular biology as an information science did not materialize until the period of the second contract for GenBank. As described above, BB&N was awarded the first five-year contract to manage GenBank. The contract was up for renewal in 1987, and on the basis of its track record in managing BIONET, IntelliGenetics submitted a proposal to manage GenBank. GenBank users had become dissatisfied with the serious delay in sequence data publication. GenBank was two years behind in disseminating sequence data it had received (Douglas Brutlag, personal communication, 19 June 1999). At a meeting in Los Alamos in 1986, Walter Goad noted that GenBank had twelve million base pairs. Other sequence collections available to researchers contained fourteen to fifteen million base pairs, so that GenBank was at least 14 to 20 percent out of date (Boswell, 1987). Concerned that researchers would turn to other, more up-to-date data sources, the NIH listed encouraging use as one of the issues they wanted IntelliGenetics to address in their proposal to manage GenBank (Duke, 1987).

IntelliGenetics proposed to solve this problem by automating the submission of gene and protein sequences. The standard method up to that time required an employee at GenBank to search the published scientific literature laboriously for sequence data, rekey these into a GenBank standard electronic format, and check them for accuracy. IntelliGenetics would automate the submission procedure with an online submission program, XGENPUB (later called "AUTHORIN").

In fact, IntelliGenetics was already progressing toward automating all levels of sequence entry and (as much as possible) analysis. As early as 1986 IntelliGenetics included SEQIN in PC/GENE, its commercial software package designed for microcomputers. SEQIN was designed for entering and editing nucleic acid sequences, and it already had the functionality needed to deposit sequences with GenBank or EMBL electronically ("PC/Gene," 1986). Transferring this program to the mainframe was a straightforward move. Indeed the online entry of original sequence data was already a feature of BIONET, since large numbers of researchers were using the IntelliGenetics GEL program on the BIONET computer. GEL was a program that accepted and analyzed data produced by all the popular sequencing methods. It provided comprehensive recordkeeping and analysis for an entire sequencing project from start to finish. The final product of the GEL program was a sequence file suitable for analysis by other programs, such as SEQ.XGENPUB, extended to this capability by allowing the scientist to annotate a sequence according to the standard GenBank format and mail the sequence and its annotation to GenBank electronically. The interface was a forms-oriented display editor that would automatically insert the sequence in the appropriate place in the form by copying the sequence from a designated file on the BIONET computer. When completed, it could be forwarded to the GenBank computer at Los Alamos; the National Institutes of Health DNA sequence library, EMBL; the nucleotide sequence database from the European Molecular Biology Laboratory; and NBRF-PIR, the National Biomedical Research Foundation's protein sequence database (Brutlag & Kristofferson, 1988).

Creating a new culture requires both carrot and stick. Making the online programs available and easy to use was one thing. Getting all molecular biologists to use them was another. In order to doubly encourage molecular biologists to comply with the new procedure of submitting their data online, the major molecular biology journals agreed to require evidence that data had been so submitted before they would consider a manuscript for review. *Nucleic Acids Research* was the first journal to enforce this transition to electronic data submission (Brutlag & Kristofferson, 1988). With these new policies and networks in place, BIONET was able to reduce the time from submission to publication and distribution of new sequence data from two years to twenty-four hours. As noted above, just a few years earlier, at the beginning of BIONET, there were only ten million base pairs published, and these had been the result of several years' effort. The new electronic submission of data generated ten million base pairs a month (Douglas Brutlag, personal communication, 19 June 1999; "Nomination for Smithsonian-ComputerWorld Award," n.d.). Walter Gilbert may have angered some of his colleagues at the 1987 Los Alamos Workshop on Automation in Decoding the Human Genome when he stated that "Sequencing the human genome is not science, it is production" (Boswell, 1987). But he surely had his finger on the pulse of the new biology.

The Matrix of Biology

The explosion of data on all levels of the biological continuum made possible by the new biotechnologies and represented powerfully by organizations such as BIONET was a source of both exhilaration and anxiety. Of primary concern to many biologists was how best to organize this massive outpouring of data in a way that would lead to deeper theoretical insight, perhaps even a unified theoretical perspective for biology. The National Institutes of Health were among those most concerned about these issues, and they organized a series of workshops to consider the new perspectives emerging from recent developments. The meetings culminated in a report from a committee chaired by Harold Morowitz titled *Models for Biomedical Research: A New Perspective* (1985). The committee foresaw the emergence of a new theoretical biology "different from theoretical physics, which consists of a small number of postulates and the procedures and apparatus for deriving predictions from those postulates." The new biology was far more than just a collection of experimental observations. Rather it was a vast array of information gaining coherence through organization into a conceptual matrix (Morowitz, 1985, p. 21). A point in the history of biology had been reached where new generalizations and higher-order biological laws were being approached but obscured by the simple mass of data and volume of literature. To move toward this new theoretical biology, the committee proposed a multidimensional matrix of biological knowledge:

> That is the complete data base of published biological experiments structured by the laws, empirical generalizations, and physical foundations of biology and connected by all the interspecific transfers of information. The matrix includes but is more than the computerized data base of biological literature, since the search methods and key words used in gaining access to that base are themselves related to the generalizations and ideas about the structure of biological knowledge. (Morowitz, 1985, p. 65)

New disciplinary requirements were imposed on the biologist who wanted to interpret and use the matrix of biological knowledge:

> The development of the matrix and the extraction of biological generalizations from it are going to require a new kind of scientist, a person familiar enough with the subject being studied to read the literature critically, yet expert enough in information science to be innovative in developing methods of classification and search. This implies the development of a new kind of theory geared explicitly to biology with its particular theory structure. It will be tied to the use of computers, which will be required to deal with the vast amount and complexity of the information, but it will be designed to search for general laws and structures that will make general biology much more easily accessible to the biomedical scientist. (Morowitz, 1985, p. 67)

Similar concerns about managing the explosion of new information motivated the Board of Regents of the National Library of Medicine. In its Long Range Plan of 1987 the NLM drew directly on the notion of the matrix of biological knowledge and elaborated upon it explicitly in terms of fashioning the new biology as an information science (Board of Regents, 1987). The Long Range Plan contained a series of recommendations that were the outcome of studies done by five different panels, including a panel that considered issues connected with building factual databases, such as sequence databases.

In the view of the panel the field of molecular biology was opening the door to an era of unprecedented understanding and control of life processes, including "automated methods now available to analyze and modify biologically important macromolecules" (Board of Regents, 1987, p. 26). The report characterized biomedical databases as representing the universal hierarchy of biological nature: cells, chromosomes, genes, proteins. Factual databases were being developed at all levels of the hierarchy, from cells to base-pair sequences. Because of the complexity of biological systems, basic research in the life sciences was increasingly dependent on automated tools to store and manipulate the large bodies of data describing the structure and function of important macromolecules. The NIH Long Range Plan stated, however, that the critical questions being asked could often only be answered by relating one biological level to another, but methods for automatically suggesting links across levels were nonexistent (Board of Regents, 1987, pp. 26–27).

> A singular and immediate window of opportunity exists for the Library in the area of molecular biology information. Because of new automated laboratory methods, genetic and biochemical data are accumulating far faster than they can be assimilated into the scientific literature. The problems of scientific research in biotechnology are increasingly problems of information science. By applying its expertise in computer technologies to the work of understanding the structure and function of living cells on a molecular level, NLM can assist and hasten the Nation's entry into a remarkable new age of knowledge in the biological sciences. (Board of Regents, 1987, p. 29)

To support and promote the entry into the new age of biological knowledge, the NIH recommended building a National Center for Biotechnology Information to serve as a repository and distribution center for this growing body of knowledge and as a laboratory for developing new information analysis and communications tools essential to the advance of the field. The proposal recommended $12.75 million per year for 1988–1990, with an additional $10 million per year for work in medical informatics (Board of Regents, 1987, pp. 46–47). The program would emphasize collaboration between computer and information scientists and biomedical researchers. In addition the NIH would support research in the areas of molecular biology database representation, retrieval-linkages, and modeling systems, while examining interfaces based on algorithms, graphics, and expert systems. The recommendation also called for the construction of online data delivery through linked regional centers and distributed database subsets.

Brave New Theory

Two different styles of work have characterized the field of molecular biology. The biophysical approach has sought to predict the function of a molecule from its structure. The biochemical approach, on the other hand, has been concerned with predicting phenotype from biochemical function. If there has been a unifying framework for the field, at least from its early days up through the 1980s, it was provided by the "central dogma" emerging from the work of James Watson, Francis Crick, Monod, and Jacob in the late 1960s, schematized as follows:

$$\text{DNA} \rightarrow \text{RNA} \rightarrow \text{Protein} \rightarrow \text{Function}$$

In this paper I have singled out molecular biologists whose Holy Grail has always been to construct a mathematized, predictive biological theory. In terms of the "central dogma" the measure of success in the enterprise of making biology predictive would be—and has been since the days of Claude Bernard—rational medicine. If one had a complete grasp of all the levels from DNA to behavioral function, including the processes of translation at each level, then one could target specific proteins or biochemical processes that may be malfunctioning and design drugs specifically to repair these disorders. For those molecular biologists with high theory ambitions, the preferred path toward achieving this goal has been based on the notion that the function of a molecule is determined by its three-dimensional folding and that the structure of proteins is uniquely contained in the linear sequence of their amino acids (Anfinsen, 1973). But determination of protein structure and function is only part of the problem confronting a theoretical biology. A fully fledged theoretical biology would want to be able to determine the biochemical function of the protein structure as well as its expected behavioral contribution within the organism. Thus biochemists have resisted the road of high theory and have pursued a solidly experimental approach aimed at eliciting common models of biochemical function across a range of mid-level biological structures from proteins and enzymes through cells. Their approach has been to identify a gene by some direct experimental procedure—determined by some property of its product or otherwise related to its phenotype—to clone it, to sequence it, to make its product, and to continue to work experimentally so as to seek an understanding of its function. This model, as Walter Gilbert has observed, was suited to "small science," experimental science conducted in a single lab (Gilbert, 1991, p. 99).

The emergence of organizations like the Brookhaven Protein Data Bank in 1971, GenBank in 1982, and BIONET in 1984, and the massive amount of sequencing data that began to become available in university and company databases, and more recently publicly through the Human Genome Initiative, has complicated this picture immensely through an unprecedented influx of new data. In the process a paradigm shift has occurred in both the intellectual and institutional structures of biology. According to some of the central players in this transformation, at the core is biology's switch from having been an observational science, limited primarily by the ability to make observations, to being a data-bound science limited by its practitioners' ability to understand large amounts of *information* derived from observations. To understand the data, the tools of information science have not only become necessary handmaidens to theory; they have also fundamentally changed the picture of biological theory itself. A new picture of theory radically different from even the biophysicists' model of theory has come into view. In terms of discipline biology has become an information science. Institutionally, it is becoming "Big Science." Gilbert characterizes the situation sharply:

> To use this flood of knowledge, which will pour across the computer networks of the world, biologists not only must become computer-literate, but also change their approach to the problem of understanding life.
>
> The next tenfold increase in the amount of information in the databases will divide the world into haves and have-nots, unless each of us connects to that information and learns how to sift through it for the parts we need. (Gilbert, 1991)

The new data-bound biology implied in Gilbert's scenario is genomics. The theoretical component of genomics might be termed *computational biology*, while its instrumental and experimental component might be considered *bioinformatics*. The fundamental dogma of this new biology, as characterized by Douglas Brutlag, reformulates the central dogma of Jacob-Monod in terms of "information flow" (Brutlag, 1994):

Genetic information → Molecular structure → Biochemical function → Biologic behavior

Walter Gilbert describes the newly forming genomic view of biology:

> The new paradigm now emerging is that all the "genes" will be known (in the sense of being resident in databases available electronically), and that the starting point of a biological investigation will be theoretical. An individual scientist will begin with a theoretical conjecture, only then turning to experiment to follow or test that hypothesis. The actual biology will continue to be done as "small science"—depending on individual insight and inspiration to produce new knowledge—but the reagents that the scientist uses will include a knowledge of the primary sequence of the organism, together with a list of all previous deductions from that sequence. (Gilbert, 1991, p. 99)

Genomics, computational biology, and bioinformatics restructure the playing field of biology, bringing a substantially modified toolkit to the repertoire of molecular biology skills developed in the 1970s. Along with the biochemistry components, new skills are now required, including machine learning, robotics, databases, statistics and probability, artificial intelligence, information theory, algorithms, and graph theory (Douglas Brutlag, personal communication).

Proclamations of the sort made by Gilbert and other promoters of genomics may seem like hyperbole. But the Human Genome Initiative and the information technology that enables it have fundamentally changed molecular biology, and indeed, may suggest similar changes in store for other domains of science. The online DNA and protein databases that I have described have not just been repositories of information for insertion into the routine work of molecular biology, and the software programs discussed in connection with IntelliGenetics and GenBank are more than retrieval aids for transporting that information back to the lab. As a set of final reflections, I want to look in more detail at some ways this software has been used to address the problems of molecular biology in order to gain a sense of the changes taking place.

Biology in Silico

To appreciate the relationship between genomics and earlier work in molecular biology, it is useful to compare approaches to the determination of structure and function. Rather than an approach deriving structure and function from first principles of the dynamics of protein folding, the bioinformatics approach involves comparing new sequences with preexisting ones and discovering structure and function by homology to known structures. This approach examines the kinds of amino acid sequences or patterns of amino acids found in each of the known protein structures. The sequences of proteins whose structure have already been determined and are already on file in the PDP are examined to infer rules or patterns applicable to novel protein sequences to predict their structure. For instance, certain amino acids, such as leucine and alanine, are very common in α-helical regions of proteins, whereas other amino acids, such as proline, are rarely if ever found in α-helices. Using patterns of amino acids or rules based on these patterns, the genome scientist can attempt to predict where helical regions will occur in proteins whose structure is unknown and for which a complete sequence exists. Clearly the lineage in this approach is work on automated learning first begun in DENDRAL and carried forward in other AI projects related to molecular biology such as MOLGEN.

The great challenge in the study of protein structure has been to predict the fold of a protein segment from its amino acid sequence. Before the advent of sequencing technology it was generally assumed that each unique protein sequence would produce a three-dimensional structure radically different from every other protein. But the new technology revealed that protein sequences are highly redundant: Only a small percentage of the total sequence is crucial to the structure and function of the protein. Moreover, while similar protein sequences generally indicate similarly folded conformations and functions, the converse does not hold. In some proteins, such as the nucleotide-binding proteins, the structural features encoding a common function are conserved, while primary sequence similarity is almost nonexistent (Rossman, Moras, & Olsen, 1974; Creighton, 1983; Birktoft & Banaszak, 1984). Methods that detect similarities solely at the primary sequence level turned out to have difficulty addressing functional associations in such sequences. A number of features often only implicit in the protein's linear or primary sequence of twenty possible amino acids turned out to be important in determining structure and function.

Such findings implied the need for more sophisticated techniques of searching than simply finding identical matches between sequences in order to elicit information about similarities between higher-ordered structures such as folds. One solution adopted early on by programs such as SEQ was to assume that if two DNA segments are evolutionarily related, their sequences will probably be related in structure and function. The related descendants are identifiable as homologues. For instance, there are more than 650 globin sequences (as in myoglobin or hemoglobin) in the protein sequence databases, all of them very similar in structure. These sequences are assumed to be related by evolutionary descent rather than having been created de novo. Many programs for searching sequence databases have been written, including an important early method written in 1970 by S. B. Needleman and C. D. Wunsch and incorporated into SEQ for aligning sequences based on homologies (Needleman & Wunsch, 1970). The method of homology depends upon assumptions related to the genetic events that could have occurred in the divergent (or convergent) evolution of proteins; namely, that homologous proteins are the result of gene duplication and subsequent mutations. If one assumes that after the duplication point mutations occur at a constant or variable rate, but randomly along the genes of the two proteins, then after a relatively short period of time the protein pairs will have nearly identical sequences. Later there will be gaps in the shared sets of base-pairs between the two proteins. Needleman and Wunsch determined the degree of homology between protein pairs by counting the number of non-identical pairs (amino acid replacements) in the homologous comparison and using this number as a measure of evolutionary distance between the amino acid sequences. A second approach was to count the minimum number of mutations represented by the non-identical pairs.

Another example of a key tool used in determining structure-function relationship is a search for sequences that correspond to small conserved regions of proteins, modular structures known as motifs. Since insertions and deletions (gaps) within a motif are not easily handled from a mathematical point of view, a more technical term, "alignment block," has been introduced that refers to conserved parts of multiple alignments containing no insertions or deletions (Bork & Gibson, 1996).

Several different kinds of motifs are related to secondary and tertiary structure. Protein scientists distinguish among four hierarchical levels of structure. Primary structure is the specific linear sequence of the twenty possible amino acids making up the building blocks of the protein. Secondary structure consists of patterns of repeating polypeptide structure within an α-helix, β-sheet, and reverse turns. Supersecondary structure refers to a few common motifs of interconnected elements of secondary structure. Segments of α-helix and β-strand often combine in specific structural motifs. One example is the α-helix-turn-helix motif found in DNA-binding proteins. This motif contains twenty-two amino acids in length that enable it to bind to DNA. Another motif at the supersecondary level is known as the Rossmann fold, in which three α-helices alternate with three parallel β-strands. This has turned out to be a general fold for binding mono- or dinucleotides and is the most common fold observed in globular proteins (Richardson & Richardson, 1989).

A higher order of modular structure is found at the tertiary level. Tertiary structure is the overall spatial arrangement of the polypeptide chain into a globular mass of hydrophobic side chains forming the central core, from which water is excluded, and more polar side chains favoring the solvent-exposed surface. Within tertiary structures are certain domains on the order of a hundred amino acids, which are themselves structural motifs. Domain motifs have been shown to be encoded by exons, individual DNA sequences that are directly translated into peptide sequences. Assuming that all contemporary proteins have been derived from a small number of original ones, Walter Gilbert and colleagues have argued that the total number of exons from which all existing protein domains have been derived is somewhere between one thousand and seven thousand (Dorit, Schoenback, & Gilbert, 1990).

Motifs are powerful tools for searching databases of known structure and function to determine the structure and function of an unknown gene or protein. The motif can serve as a kind of probe for searching the database or some new sequence, testing for the presence of that motif. The PROCITE database, for example, has more than a thousand of these motifs (Bairoch, 1991). With such a library of motifs one can take a new sequence and use each one of the motifs to get clues about its structure. Suppose, for example, the sequence of a gene or protein has been determined. Then the most common way to investigate its biologic function is simply to compare its sequence with all known DNA or protein sequences in the databases and note any strong similarities. The particular gene or protein that has just been determined will of course not be found in the databases, but a homologue from another organism or a gene or protein having a related function may be found. The evolutionary similarity implies a common ancestor and hence a common function. Searching with motif probes refines the determination of the fold regions of the protein. These methods become more and more successful as the databases grow larger and as the sensitivity of the search procedure increases. Bork, Ouzounis, and Sander (1994) state that the likelihood of identifying homologues is currently higher than 80 percent for bacteria, 70 percent for yeast, and about 60 percent for animal sequence series (Bork & Gibson, 1996).

The all-or-nothing character of consensus sequences—a sequence either matches or it does not—led researchers to modify this technique to introduce degrees of similarity among aligned sequences as a way of detecting similarities between proteins, even distantly related ones. Knowing the function of a protein in some genome, such as *E. coli*, for instance, might suggest the same function of a closely related protein in an animal or human genome (Patthy, 1996). Moreover, as noted above, different amino acids can fit the same pattern, such as the helix-turn-helix, so that a representation of sequence pattern in which alternative amino acids are acceptable, as well as regions in which a variable number of amino acids may occur, are desirable ways of extending the power of straightforward consensus sequence comparison. One such technique is to use weights or frequencies to specify greater tolerance in some positions than in others. An illustration of the success of this approach is provided by the DNA-binding proteins mentioned above, which contain a helix-turn-helix motif twenty-two acids in length (Brennan & Mathews, 1989). Comparison of the linear amino acid sequences of these proteins revealed no consensus sequence that could distinguish them from any other protein. A weight matrix is constructed by determining the frequency with which each amino acid appears at each position, and then converting these numbers to a measure of the probability of occurrence of each acid. This weight matrix can be applied to measure the likelihood that any given sequence twenty-two amino acids long is related to the helix-turn-helix family. A further modification of the weight matrix is the profile, which allows one to estimate the probability that any amino acid will appear in a specific position (Gribskov et al., 1987; Gribskov et al., 1988).

In addition to consensus sequences, weight matrices, and profiles, a further class of strategies for determining structure-function relations are various sequence alignment methods. In order to detect homologies between distantly related proteins, one method is to assign a measure of similarity to each pair of amino acids, and then add up these pairwise scores for the entire alignment (Schwartz & Dayhoff, 1979). Related proteins will not have identical amino acids aligned, but they do have chemically similar or replaceable amino acids in similar positions. In a scoring method developed by R. M. Schwartz and M. O. Dayhoff, for example, amino acid pairs that are identical or chemically similar were given positive scores, and pairs of amino acids that are not related were assigned negative similarity scores.

A dramatic illustration of how sequence alignment tools can be brought to bear on determining function and structure is provided by the case of cystic fibrosis. Cystic fibrosis is caused by aberrant regulation of chloride transport across epithelial cells in the pulmonary tree, the intestine, the exocrine pancreas, and apocrine sweat glands. This disorder was identified as being caused by defects in the cystic fibrosis transmembrane conductance regulator protein (CFTR). After the CFTR gene was isolated in 1989, its protein product was identified as producing a chloride channel, which depends for its activity on the phosphorylation of particular residues within the regulatory region of the protein. Using computer-based sequence alignment tools of the sort described above, it was established that a consensus sequence for nucleotide binding folds that bind ATP are present near the regulatory region and that 70 percent of cystic fibrosis mutations are accounted for by a three base-pair deletion that removes a phenylalanine residue within the first nucleotide-binding position. A significant portion of the remainder of cystic fibrosis mutations affect a second nucleotide-binding domain near the regulatory region (Hyde et al., 1990; Kerem et al., 1989; Kerem et al., 1990; Riordan et al., 1989).

In working out the folds and binding domains for the CFTR protein, S. C. Hyde, P. Emsley, M. J. Hartshorn, and colleagues (1990) used sequence alignment methods similar to those available in early models of the IntelliGenetics software suite. They used the Chou-Fasman algorithm (1973) for identifying consensus sequences and the Quanta modeling package produced by Polygen Corporation (Waltham, Massachusetts) for modeling the protein and its binding sites (Hyde et al., 1990). In 1992 IntelliGenetics introduced BLAZE, an even more rapid search program running on a massively parallel computer. As an example of how computational genomics can be used to solve structure-function problems in molecular biology, Brutlag repeated the CFTR case using BLAZE (Brutlag, 1994). A sequence similarity search compared the CFTR protein to more than twenty-six thousand proteins in a protein database of more than nine million residues, resulting in a list of twenty-seven top similar proteins, all of which strongly suggested the CFTR protein is a membrane protein involved in secretion. Another feature of the comparison result was that significant homologies were shown with ATP-binding transport proteins, further strengthening the identification of CFTR as a membrane protein. The search algorithm identified two consensus sequence motifs in the protein sequence of the cystic fibrosis gene product that corresponded to the two sites on the protein involved in binding nucleotides. The search also turned up distant homologies between the CFTR protein and proteins in *E. coli* and yeast. The entire search took three hours. Such examples offer convincing evidence that tools of computational molecular biology can lead to the understanding of protein function.

The methods for analyzing sequence data discussed above were just the beginnings of an explosion of database mining tools for genomics that is continuing to take place.[3] In the process biology is becoming even more aptly characterized as an information science (Hughes et al., 1999; IntelliGenetics & MasPar Computer Corporation, 1992). Advances in the field have led to large-scale automation of sequencing in genome centers employing robots. The success this large-scale sequencing of genes has enjoyed has in turn spawned a similar approach to applying automation to sequencing proteins, a new area complementary to genomics called proteomics. Similar in concept to genomics, which seeks to identify all genes, proteomics aims to develop techniques that can rapidly identify the type, amount, and activities of the thousands of proteins in a cell. Indeed, new biotechnology companies have started marketing technologies and services for mining protein information en masse. Oxford Glycosciences (OGS) in Abingdon, England, has automated the laborious technique of two-dimensional gel electrophoresis.[4] In the OGS process an electric current applied to a sample on a polymer gel separates the proteins, first by their unique electric charge characteristics and then by size. A dye attaches to each separated protein arrayed across the gel. Then a digital imaging device automatically detects protein levels by how much the dye fluoresces. Each of the five thousand to six thousand proteins that may be assayed in a sample in the course of a few days is channeled through a mass spectrometer that determines its amino acid sequence. The identity of a protein can be determined by comparing the amino acid sequence with information contained in numerous gene and protein databases. One imaged array of proteins can be contrasted with another to find proteins specific to a disease.

In order to keep pace with this flood of data emerging from automated sequencing, genome researchers have in turn looked increasingly to artificial intelligence, machine learning, and even robotics in developing automated methods for discovering patterns and protein motifs from sequence data. The power of these methods is their ability both to represent structural features rather than strictly evolutionary steps and to discover motifs from sequences automatically. The methods developed in the field of machine learning have been used to extract conserved residues, discover pairs of correlated resi-

[3] See, for instance, the National Institute of General Medical Science (NIGMS) "Protein Structure Initiative Meeting Summary," 24 April 1998 at http://www.nih.gov/nigms/news/reports/protein_structure.html.

[4] See the discussion of this technology at the site of Oxford Glycosciences: http://www.ogs.com/proteome/home.html.

dues, and find higher-order relationships between residues as well. Techniques from the field of machine learning have included perceptrons, discriminant analysis, neural networks, Bayesian networks, hidden Markov models, minimal length encoding, and context-free grammars (Hunter, 1993). Important methods for evaluating and validating novel protein motifs have also derived from the machine learning area.

An example of this effort to scale up and automate the discovery of structure and function is EMOTIF (for "electronic-motif"), a program for discovering conserved sequence motifs from families of aligned protein sequences developed by the Brutlag Bioinformatics Group at Stanford (Nevill-Manning et al., 1998).[5] Protein sequence motifs are usually generated manually with a single "best" motif optimized at one level of specificity and sensitivity. Brutlag's aim was to automate this procedure. An automated method requires knowledge about sequence conservation. For EMOTIF, this knowledge is encoded as a particular allowed set of amino acid substitution groups. Given an aligned set of protein sequences, EMOTIF works by generating a set of motifs with a wide range of specificities and sensitivities. EMOTIF can also generate motifs that describe possible subfamilies of a protein superfamily. The EMOTIF program works by generating a new database, called IDENTIFY, of fifty thousand motifs from the combined seven thousand protein alignments in two widely used public databases, the PRINTS and BLOCKS databases. By changing the set of substitution groups, the algorithm can be adapted for generating entirely new sets of motifs.

Highly specific motifs are well suited for searching entire proteomes. IDENTIFY assigns biological functions to proteins based on links between each motif and the BLOCKS or PRINTS databases that describe the family of proteins from which it was derived. Because these protein families typically have several members, a match to a motif may provide an association with several other members of the family. In addition, when a match to a motif is obtained, that motif may be used to search sequence databases, such as SWISS-PROT and GenPept, for other proteins that share this motif. In their paper introducing these new programs C. G. Nevill-Manning, T. D. Wu, and Brutlag showed that EMOTIF and IDENTIFY successfully assigned functions automatically to 25 to 30 percent of the proteins in several bacterial genomes and automatically assigned functions to 172 proteins of previously unknown function in the yeast genome.

Many molecular biologists who welcomed the Human Genome Initiative with open arms undoubtedly believed that when the genome was sequenced everyone would return to the lab to conduct their experiments in a business-as-usual fashion, empowered with a richer set of fundamental data. The developments in automation, the resulting explosion of data, and the introduction of tools of information science to master this data have changed the playing field forever: There may be no "lab" to return to. In its place is a workstation hooked to a massively parallel computer, producing simulations by drawing on the data streams of the major databanks, and carrying out "experiments" *in silico* rather than in vitro. The result of biology's metamorphosis into an information science just may be the relocation of the lab to the industrial park and the dustbin of history.

[5] EMOTIF can be viewed at http://motif.stanford.edu/emotif.

References

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science, 181*(4096), 223–230.

Bairoch, A. PROSITE: A dictionary of sites and patterns in proteins. *Nucleic Acids Research, 19,* 2241.

Bernstein, F. C., Koetzle, T. F., et al. (1977). The Protein Data Bank: A computer based archival file for macromolecular structure. *Journal of Molecular Biology, 112,* 535–542.

BIONET users status. (1986, April 3 and October 9). From BIONET managers' meetings. (Stanford University Special Collections, Brutlag Papers, Fldr BIONET). Stanford, CA: Stanford University.

Birktoft, J. J., & Banaszak, L. J. (1984). Structure-function relationships among nicotinamide-adenine dinucleotide dependent oxidoreductases. In M. T. W. Hearn (Ed.), *Peptide and Protein Reviews* (Vol. 4, pp. 1–47). New York: Marcel Dekker.

Board of Regents. (1987). *NLM long range plan (report of the Board of Regents).* Bethesda, MD: National Library of Medicine.

Bork, P., & Gibson, T. J. (1996). Applying motif and profile searches. In R. F. Doolittle (Ed.), *Computer Methods for Macromolecular Sequence Analysis* (Vol. 266 in Methods in Enzymology, pp. 162–184, especially p. 163). San Diego: Academic Press.

Bork, P., Ouzounis, C., & Sander, C. (1994). From genome sequences to protein function. *Current Opinions in Structural Biology, 4*(3), 393–403.

Boswell, S. (January 9, 1987). Los Alamos workshop—Exploring the role of robotics and automation in decoding the human genome. (IntelliGenetics trip report. In Stanford Special Collections, Brutlag Papers, Fldr BIONET). Stanford, CA: Stanford University.

Brennan, R. G., & Mathews, B. W. (1989). The helix-turn-helix binding motif. *Journal of Biological Chemistry, 264,* 1903.

Brutlag, D. L. (1994). Understanding the human genome. In P. Leder, D. A. Clayton, & E. Rubenstein (Eds.), Scientific American: *Introduction to Molecular Medicine* (pp. 153–168). New York: Scientific American, Inc.

Brutlag, D. L., & Kristofferson, D. (1988). BIONET: An NIH computer resource for molecular biology. In R. R. Colwell (Ed.), *Biomolecular data: A resource in transition* (pp. 287–294). Oxford: Oxford University Press.

Business plan for IntelliGenetics. (1981, May 8). (Stanford Special Collections, Brutlag Papers, Fldr IntelliGenetics, p. 2). Stanford, CA: Stanford University.

Creighton, T. E. (1983). *Proteins: Structure and molecular properties.* New York: W. H. Freeman.

Dorit, R. L., Schoenback, L., & Gilbert, W. (1990). How big is the universe of exons? *Science, 250,* 1377.

Doyle, R. (1997). *On beyond living: Rhetorical transformations of the life sciences.* Stanford, CA: Stanford University Press.

Duke, B. H., Contracting Officer, NIH, to IntelliGenetics, Inc. (1987, June 3). Request for revised proposal in response to request for proposals RFP no. NIH-GM-87-04 titled "Nucleic Acid Sequence Data Bank." Letter with attachment. (Stanford Special Collections, Brutlag Papers, Fldr BIONET). Stanford, CA: Stanford University.

Feigenbaum, E. A., Buchanan, B., et al. (April 1980). A proposal for continuation of the MOLGEN project: A computer science application to molecular biology (Computer Science Department, Stanford University, Heuristic Programming Project, Technical Report No. HPP-80-5, Section 1), p. 1. Stanford, CA: Stanford University.

Feigenbaum, E. A., & Martin, N. (1977). Proposal: MOLGEN—a computer science application to molecular genetics (Heuristic Programming Project, Stanford University, Technical Report No: HPP-78-18,1977). Stanford, CA: Stanford University.

Friedhoff, R. M., & Benzon, W. (1989) *The second computer revolution: Visualization.* New York: W. H. Freeman.

Friedland, P. (1979). *Knowledge-based experiment design in molecular genetics.* Unpublished doctoral dissertation, Stanford University.

Friedland, P. (1984, April 27). BIONET organizational plans. Company confidential memo. (Stanford University Special Collections, Brutlag Papers, Fldr BIONET, p. 1). Stanford, CA: Stanford Universisty.

Friedland, P., Brutlag, D. L., Clayton, J., & Kedes, L. H. (1982). SEQ: A nucleotide sequence analysis and recombinant system. *Nucleic Acids Research, 10,* 279–294.

Friedland, P., to Reimers, N. Subject: Software licensing agreement: April 2, 1984. (Stanford University Special Collections, Fldr IntelliGenetics). Stanford, CA: Stanford University.

Gilbert, W. (1991). Towards a paradigm shift in biology. *Nature, 349,* 99.

Gribskov, M., Homyak, M., et al. (1988). Profile scanning for three-dimensional structural patterns in protein sequences. *Computer Applications in the Biosciences, 4,* 61.

Gribskov, M., McLachlan, A. D., et al. (1987). Profile analysis: Detection of distantly related proteins. *Proceedings of the National Academy of Sciences, 84,* 4355.

Hall, S. S. (1995). Protein images update natural history. *Science, 267*(3 February), 620–624.

Holbrook, S. R., Muskal, S. M., et al. (1993). Predicting protein structural features with artificial neural networks. In L. Hunter (Ed.), *Artificial intelligence and molecular biology* (pp. 161–194). Menlo Park, CA: AAAI (American Association for Artificial Intelligence) Press.

Huberman, J. (1989). BIONET: Computer power for the rest of us. (Stanford Special Collections, Brutlag Papers, Fldr BIONET). Stanford, CA: Stanford University.

Hughes, T. P., et al. (Eds.). (1999). *Funding a revolution: Government support for computing research.* Washington, DC: National Academy Press.

Hunter, L. (Ed.). (1993). *Artificial intelligence and molecular biology.* Menlo Park, CA: AAAI Press.

Hyde, S. C., Emsley, P., et al. (1990). Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. *Nature, 346,* 362–365.

IntelliGenetics. (1987). *Introduction to BIONET: A computer resource for molecular biology. User manual for Bionet subscribers, Release 2.3.* Mountain View, CA: IntelliGenetics.

Katz, L., & Levinthal, C. (1972). Interactive computer graphics and the representation of complex biological structures. *Annual Reviews in Biophysics and Bioengineering, 1,* 465–504.

Kerem, B. S., Rommens, J. M., et al. (1989). Identification of the cystic fibrosis gene: Genetic analysis. *Science, 245,* 1073–1080.

Kerem, B. S., Zielenski, J., et al. (1990). Identification of mutations in regions corresponding to the two putative nucleotide (ATP)-binding folds of the cystic fibrosis gene. *Proceedings of the National Academy of Sciences, 87,* 8447–8451.

Korn, L. J., Queen, C. L., & Wegman, M. N. (1977). Computer analysis of nucleic acid regulatory sequences. *Proceedings of the National Academy of Sciences, 74,* 4516–4520.

Lederberg, J. (n.d.). How DENDRAL was conceived and born (Stanford Technical Reports 048087-54, Knowledge Systems Laboratory Report No. KSL 87-54). Stanford, CA: Stanford University.

Lederberg, J. (1969). Topology of molecules. In National Research Council Committee on Support of Research in the Mathematical Sciences (Ed.), *The mathematical sciences: A collection of essays* (pp. 37–51). Cambridge, MA: MIT Press.

Lederberg, J., Sutherland, G. L., Buchanan, B. G., & Feigenbaum, E. (1969 November). A heuristic program for solving a scientific inference problem: Summary of motivation and implementation (Stanford Technical Reports 026104, Stanford Artificial Intelligence Project Memo AIM-104). Stanford, CA: Stanford University.

Ledley, R. S. (1965). *Use of computers in biology and medicine.* New York: McGraw-Hill.

Levinthal, C. (1966). Molecular model-building by computer. *Scientific American, 214*(6), 42–52.

Levinthal, C., Barry, C. D., Ward, S. A., & Zwick, M. (1968). Computer graphics in macromolecular chemistry. In D. Secrest & J. Nievergelt (Eds.), *Emerging concepts in computer graphics* (pp. 231–253). New York: W. A. Benjamin.

Lewin, R. (1984). National networks for molecular biologists. *Science, 223,* 1379–1380.

Maxam, A. M., to GENET Community. (1982, August 23). Subject: Closing of GENET. (Stanford University Special Collections, Peter Friedland Papers, Fldr GENET). Stanford, CA: Stanford University.

McCormick, B. H., DeFanti, T. A., & Brown, M. D. (1987). Visualization in Scientific Computing. NSF Report. *Computer Graphics, 21*(6), special issue.

Minutes of the meeting of the National Advisory Committee for BIONET. (1985, March 23). (Final version prepared 1985, August 1). (Stanford University Special Collections, Brutlag Papers, Fldr BIONET, p. 4). Stanford, CA: Stanford University.

Morowitz, H. (1985). *Models for biomedical research: A new perspective.* Washington, DC: National Academy of Sciences Press.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology, 48,* 443.

Nevill-Manning, C. G., Wu, T. D., et al. (1998). Highly specific protein sequence motifs for genomic analysis. *Proceedings of the National Academy of Sciences, USA, 95*(11), 5865–5871.

NIH Special Study Section. (March 17–19, 1983). BIONET, national computer resource for molecular biology. (Stanford University Special Collections, Brutlag Papers, p. 2). Stanford, CA: Stanford University.

Nomination for Smithsonian-Computerworld Award. (n.d.). (Stanford Special Collections, Brutlag Papers, Fldr Smithsonian Computerworld Award). Stanford, CA: Stanford University.

Oettinger, A. G. (1966). The uses of computers in science. *Scientific American, 215*(3), 161–172.

Panel on Information Technology and the Conduct of Research, National Academy of Sciences. *Information technology and the conduct of research: The user's view.* (1989). Washington, DC: National Academy of Sciences.

Patthy, L. (1996). Consensus approaches in detection of distant homologies. In R. F. Doolittle (Ed.), *Computer methods for macromolecular sequence analysis* (Vol. 266 in Methods in Enzymology, pp. 184–198). San Diego: Academic Press.

PC/Gene: Microcomputer software for protein chemists and molecular biologists, user manual. (1986). Mountain View, CA: IntelliGenetics, pp. 99–120.

Ramsey, D. M. (Ed.). (1968). *Image processing in biological science.* Berkeley/Los Angeles: University of California Press.

Richardson, J. S., & Richardson, D. C. (1989). Principles and patterns of protein conformation. In G. D. Gasman (Ed.), *Prediction of protein structure and the principles of protein conformation.* New York: Plenum Press.

Rindfleisch, T., Friedland, P., & Clayton, J. (1981). The GENET guest service on SUMEX (SUMEX-AIM Report). (Stanford University Special Collections, Friedland Papers, Fldr GENET). Stanford, CA: Stanford University.

Riordan, J. R., Rommens, J. M., et al. (1989). Identification of the cystic fibrosis gene: Cloning and characterization of complementary RNA. *Science, 245,* 1066–1073.

Rossman, M. G., Moras, D., & Olsen, K. W. (1974). Chemical and biological evolution of a nucleotide-binding protein. *Nature, 250,* 194–199.

Schwartz, R. M., & Dayhoff, M. O. (1979). Matrices for detecting distant relationships. *Atlas of Protein Structure, 5*(Supplement 3), 353.

Second cut at interaction language and procedure. (n.d.). In Chemistry project. (Edward Feigenbaum Papers, Stanford Special Collections SC-340, Box 13). Stanford, CA: Stanford University.

Smith, D. H., Brutlag, D., Friedland, P., & Kedes, L. H. (1986). BIONET: National computer resource for molecular biology. *Nucleic Acids Research, 14*, 17–20.

Staden, R. (1977). Sequence data handling by computer. *Nucleic Acids Research, 4,* 4037–4051.

Staden, R. (1978). Further procedures for sequence analysis by computer. *Nucleic Acids Research, 5,* 1013–1015.

Staden, R. (1979). A strategy of DNA sequencing employing computer programs. *Nucleic Acids Research, 6,* 2602–2610.

Stefik, M. (1977). Inferring DNA structures from segmentation data. *Artificial Intelligence, 11,* 85–114.

Watson, H. C. (1969). The stereochemistry of the protein myoglobin. *Progress in Stereochemistry, 4,* 299–333.

Secret Scientific Communities: Classification and Scientific Communication in the DOE and DoD

Robert W. Seidel

Introduction

Science and secrecy have an ancient lineage, dating back to the Egyptians and Babylonians who developed number systems, geometry, as well as secret codes, and evolving through the hermetic laboratories of the high Middle Ages, where alchemy, in particular, protected knowledge of the elixir of life and the transformation of base materials into gold. Historians like Maurice Crosland (1962) who seek to decipher many of these codes are still frustrated, as are Cold War historians, in their attempts to break through the barriers of classification and security to learn about the activity of modern alchemists.

The Rosetta Stone of secret science is a security clearance and a need to know that permits access to classified information and facilities. Twenty years ago the teams of historians that prepared histories of the Lawrence Radiation Laboratory, the Los Alamos Scientific Laboratory, and Sandia Laboratories obtained such clearances and posed questions that opened drawers. Their products, still forthcoming, represent an attempt to apply a traditional approach to history—the acquisition of the requisite language with which to investigate a source—here, the new language of classification and security (Heilbron & Seidel, 1989; Furman, 1990; Hoddeson et al., 1993).

Like learning hieroglyphics or cuneiform, acquiring an understanding of classification and security is useful for historians who investigate such topics, including academic historians as well as agency historians. In the latter case, however, the final product must be checked by a "native speaker," usually a civil servant authorized to declassify information whose provenance is classified documents or research. The potential restriction of the researcher's free expression prevents most historians from applying this technique. Instead, they seek to declassify information through the Freedom of Information Act. This translation process is more difficult because it requires a knowledge of the existence of the source, seldom results in a translation that is complete, and often requires years to be completed. Nevertheless, hope springs eternal among these historians that at the stroke of a legislative pen or the bang of a judicial gavel, the walls of national security will crumble into dust, and they will be able to examine the original documents without need for translation. Their latest champion, Senator Daniel Patrick Moynihan, has proposed legislation to dismantle government secrecy and has written about it (1998). The pitfalls put in the path of this legislation have been formidable ("Update," 1998; "Administration Underscores," 1998; "President Critical," 1998).

However, millions of declassified documents would be unintelligible without a working knowledge of the original language, in which there are many false cognates. These result because information systems incorporate different levels of information: One finds not only "facts" but also structural information related to provenance and program, a wealth of acronyms for which there is no single Rosetta Stone, and many other clues that require an intimate understanding of the uses of such information.

Historians of Department of Defense (DoD) and Department of Energy (DOE) laboratories have provided a context for understanding information used within them. These "secret scientific communities" balance an interest in the dissemination of fundamental and applied scientific work and a concern about protection of information that might, if released, damage national security. Therefore, a historical understanding of the reconciliation between science and secrecy in these institutions provides a means of examining the dynamic relationship between the ideally open and the nearly closed.

Secret military research was unusual prior to World War I, although it occurred even in ancient times, Greek fire being a very early example (Long & Roland, 1994). Meanwhile, scientific information systems created in the seventeenth century have also been greatly ramified in this century (Price, 1963). This essay focuses on the last half century, not only because it saw the creation of secret scientific communities but also because it has seen the production of massive quantities of secret science, unprecedented in the history of science, as well as the evolution of new information systems to replace those of the early modern period.

In 1940 Leo Szilard feared that the discovery of fission by Otto Hahn and Fritz Strassmann at the Max Planck Institute for Chemical Research in Berlin would lead to the development of nuclear weapons. He urged his colleagues in the United States, Britain, and France to refrain from publishing research on the chain reaction in uranium. When Frédéric Joliot-Curie and his colleagues published nevertheless, this preliminary attempt at scientific self-censorship collapsed, and a flood of articles on fission chain reactions filled the scientific journals (Lanoette, 1992; Weart, 1979).

It required a higher power than Szilard's to stem the scientific passion for priority during the "phony war" of 1939–1940. The National Defense Research Committee (NDRC), which Vannevar Bush organized, supplied it (Meigs, 1982; Zachary, 1997; Stewart, 1948; Baxter, 1946). Bush had served as the vice president of MIT, president of the Carnegie Institute of Washington, and chairman of the National Advisory Committee on Aeronautics (NACA), one of the few scientific advisory boards remaining from the mobilization of World War I. He persuaded President Franklin D. Roosevelt to authorize the NDRC and its successor, the Office of Scientific Research and Development (OSRD), to oversee academic and industrial research supported by the federal government.

Bush took NACA as a model in many ways for the mobilization of science, in particular, in organizing science and technology information. NACA had institutionalized the technical report as its preferred form of scientific communication (Wooster, 1987). Industrial laboratories had developed analogous forms of internal technical communication (Hounshel & Smith, 1988; Reich, 1985). The advantages of the technical report for rapid communication as opposed to more conventional forms of scientific information are obvious. Because it is not intended for publication, it requires neither elaborate documentation nor peer review. Distribution is often limited to those who are directly concerned with the work reported, and, when classified, a technical report is accessible only to those who have the appropriate security clearances and a certified need to know its contents.

As information science pioneer Harold Wooster pointed out, "For some reason, technical documentary reports are regarded as second class citizens, which is a pity. Reports have a long and honorable history going back to 1915 and the old National Advisory Committee for Aeronautics" (Burton & Green, 1961, pp. 35–37). Bush incorporated them into the standard NDRC and OSRD contract. It called for contractors to report "the progress of such studies and investigations from time to time as requested by the Scientific Officer, and . . . furnish a complete final report of such findings and conclusions." Moreover, it laid out stringent security provisions, prohibiting the disclosure of any information concerning the contract or the results of the work to anyone except employees assigned to it during the course of the war. Failure to safeguard the information subjected "employees and contractors to criminal liability." Aliens and individuals determined by the contracting officer to be undesirable were prohibited access to contractor facilities and work (Stewart, 1948, Appendix 2).

Simultaneously, Bush continued a decade-old effort to harness microfilm as a means of information storage and retrieval. He won NDRC support for these efforts at the beginning of the war, but his efforts were frustrated by design and mechanical problems. By the end of the war he could still only project his vision of their potential (1945). He also sought to automate cryptography (Burke, 1994).

Military Security

Security classification of technical information was a relatively new process at the beginning of World War II. In 1936 Congress unanimously passed Senate Bill 1485, which authorized the president to define "vital military and naval installations or equipment as requiring protection against the general dissemination of information." Roosevelt's Executive Order 8381 gave him control of the army's and navy's classification system. The system assigned a "Secret" classification to information that could "endanger national security," or cause "serious injury to the interests or prestige of the nation, or any governmental activity thereof, and would be of great advantage to a foreign nation." A lower level of security applied to "Confidential" documents that would not endanger the national security but met the other criteria, and "Restricted" documents, which should not be published or communicated except for official purposes. The system of classification became more elaborate and restrictive during the war. The classification "Top Secret" came into use in 1944 "to cover secret documents, information, and material, the security aspect of which was paramount, and whose unauthorized disclosure would cause exceptionally grave damage to the nation." This classification severely retarded the communication of information to which it applied, according to Stewart (1948, pp. 250–251).

Bush and his deputy, Harvard President James Bryant Conant, had only had brief experience with military research in World War I. Consequently, "Secrecy as an institutional procedure also possessed for Bush and Conant none of the coercive symbolism with which it is associated today" (Meigs, 1982, p. 18; Shils, 1956, pp. 176–191). Others were more concerned. Theoretical physicist E. U. Condon, whom J. Robert Oppenheimer selected as his assistant director at Los Alamos, preferred to resign rather than administer the army's system there (Jones, 1985). Moreover, although Oppenheimer was willing to don a military uniform to get the job done, a number of the scientists whom he tried to recruit refused khaki; and so the laboratory was originally staffed with civilian scientists, with an intent—never fulfilled—of militarizing it when research reached the development phase (Bush & Conant, 1983).

Los Alamos was organized to overcome the disadvantages of classification and compartmentalization by concentrating various theoretical and experimental studies associated with the design of nuclear weapons in one place to enhance communication and increase the pace of the work. As Oppenheimer's experimental coordinator, physicist John Manley of the University of Chicago recalled,

> I had to chase around the country because there were . . . nine separate contracts with universities that had accelerators which could be used as neutron sources. . . . The problem of liaison among all the groups was a fantastically difficult one. We couldn't of course, use long distance telephone; our work was classified. Teletype connections that *could* carry classified messages were limited and next to hopeless for trying to unsnarl experimental difficulties. . . . We were so upset about the situation that shortly after General Groves was appointed . . . we approached him about establishing a new laboratory where one could bring all these separate groups, have an interchange of ideas on the experimental and theoretical difficulties instead of all this running around the country between groups of theorists and experimentalists. This consolidation was the main reason for Los Alamos. (1980)

Once together at the remote site, scientists were able to compare notes and set forward a working program of research. Manley, E. M. McMillan, and Hugh Bradner planned the experimental equipment and layout of the laboratory. The theoretical situation was set out by Robert Serber in a series of lectures in 1943 and published as LA-1, the first technical report of the laboratory (Cf. R. Serber, 1993). A review committee composed of senior scientists in the project then recommended appropriate courses of action to take.

In both MED (Manhattan Engineering District) and OSRD laboratories scientists recorded their work in technical reports. They also set up technical libraries to provide both access and security. For example, Oppenheimer recruited Robert Serber's wife, Charlotte, to run the library at Los Alamos. She was not a trained librarian, but he believed a professional librarian would be too meticulous to keep pace with the project. To assist in the work, they arranged for the loan of a number of books from the physics library at the University of California, Berkeley, and subscribed to physics journals through the university's business office, which surreptitiously transferred them to the site.

Security measures were primitive at first. Project secret reports and confidential mail were originally filed with platinum and gold foils, and scientists' cash was deposited in Oppenheimer's safe: "It had a unique combination, for although it was a three-tumbler affair, it required a swift kick at one point or it refused to open" (C. Serber, 1988, p. 65). David Hawkins, Oppenheimer's administrative director, instituted a nightly search for secret documents left unsecured. The punishments meted out were stiff fines or responsibility for a week of these security inspections. "These inspectors turned out to be the most efficient," Charlotte Serber recalled. "They seemed to get a vicious delight in discovering another offender." However, when Emilio Segrè was confronted for having left a secret document on his desk and ordered to make the rounds, he argued, "That paper, it was all wrong. I would only have confused the enemy!"(R. Serber, 1998, p. 80).

By the end of the war "the library was an odd place," Serber recounted. "It was the center for all gossip. It was a hangout. It had a document room and vault. It was the production center for all secret reports written on the Project. It was the sole owner of a ditto machine on which was run off everything from scientific reports to notices of ski club meetings, but it really was a library, too" (C. Serber, 1988, p. 70).

At the MIT Radiation Laboratory, Samuel Goudsmit organized a document room. Beginning with British reports that accompanied the transfer of radar magnetron technology in 1941, he recorded, indexed, and advertised these and other incoming reports in laboratory publications, and printed and distributed the Radiation Laboratory's own reports. It was, in the words of Henry Guerlac, "a combined reference library, editorial room, printing concern, security office, distribution center, and general information bureau" (1987, p. 677). Technical report libraries formed the neurons of the nervous system of secret scientific communities during World War II, and the Army Command Administrative Network provided secure communications between neurons, using enciphered teletypewriter messages. Bell Telephone Laboratory work on encryption provided the necessary equipment. It may also have inspired Claude Shannon's work on the theory of secret communication (Jones, 1985; Fagen, 1978; Shannon, 1993).

Postwar Secrets

The war came to an end thanks to the crucial role of radar and the definitive closure brought by the deployment of nuclear weapons to Japan. After the war, scientists hoped to return to the *status quo ante bellum* by publishing the scientific results of their work, returning to their academic and industrial laboratories, and resuming the studies that had been interrupted by OSRD and MED mobilization. To reap the scientific harvest of the war, they had to declassify wartime reports or write up their research in an unclassified form. Some scientists, like Edwin M. McMillan, who had discovered the principle of synchronous acceleration of subatomic particles while at Los Alamos, went so far as to smuggle papers out to avoid this delay (Wilson, 1993). Luis Alvarez, another missionary to Los Alamos from Lawrence's Radiation Laboratory, flew back from Hiroshima full of thoughts about a linear accelerator that he had conceived while at MIT and published without benefit of review (Goldman, 1986).

The effort to write up and declassify the wartime accomplishments was extensive. The Smyth Report led to the National Nuclear Energy Series. The Radiation Laboratory produced its own series of reports. These reports were shepherded to publication by scores of scientists, and the dispersion of scientists and engineers to universities and industry accelerated the informal dissemination of information. Like Samuel Slater, they carried in their heads the detailed plans for another industrial revolution in America.

Conant and Groves had anticipated the demand for information about the Manhattan Project. They commissioned physicist Henry Smyth of Princeton to write his famous report to provide as much information as possible, without disclosing "military secrets." Richard Tolman and his OSRD staff censored it. "Many changes in the original draft became necessary as our security criteria were applied to it," Groves (1983, pp. 348–349) remembered. "Copies of pertinent sections were given a final review by scientists in the various parts of the project, both for factual content and for security considerations. In order to speed up the process, officer couriers delivered the copies, and generally waited until the review was committed." Groves recognized the scientific and personal need that "everyone be accorded the recognition he deserved. This, we felt, would lessen the chances of future security breaks."

Groves found that scientists were not content with the Smyth Report and that they wanted to publish their work in traditional scientific journals. He appointed a Committee on Declassification, composed of the leaders of the wartime projects under his command, to advise him on the scope of declassification and the distribution of classified materials to cleared organizations and individuals. Groves ordered the study in a letter to R. C. Tolman on 2 November 1945 and appointed himself as chair and E. O. Lawrence, A. H. Compton, Harold Urey, Frank H. Spedding, R. F. Bacher, and J. Robert Oppenheimer to the committee (First report, 1945). The committee concluded that national welfare would best be served by almost total declassification, and national security would not benefit in the long term from concealing scientific information. While there were "probably good reasons for keeping close control of much scientific information if it is believed that there is a likelihood of war within the next five or ten years . . . this would weaken us disastrously for the future—perhaps twenty years hence." The committee recommended release of information that was either substantially known outside the project, was readily obtainable by theoretical or experimental work, or that would enhance American scientific or technological leadership. Information that could weaken the American military or international position would remain classified until there was "a real reduction in the threat of atomic warfare," as determined by the president of the United States and by Congress.

This recommendation left classification authority in the hands of the government and prey to the winds and rumors of war. The failure of the United Nations to internationalize nuclear power in the postwar period and the Soviet Union's development of its atomic bomb meant that complete declassification was never undertaken. There were, however, substantial amounts of material declassified in the first years after the war.

By April 1946 a Declassification Guide and Manual of Procedures had been completed and distributed throughout the MED laboratories. Groves set up a "Committee of Senior Responsible Reviewers" made up of scientists from various compartments of the project, who supervised the work of one hundred and fifteen responsible reviewers and "a considerable number of declassification officers, clerks, and typists working in the interest of the flow of scientific and technical information from restricted areas into normal channels to the maximum extent consistent with national policy and interest" (Manley, 1950, pp. 17–18). This consistency was the hobgoblin of great minds. It is impossible to estimate how much was lost to science because of the need to review and release work months, if not years, after it had been written. To be sure, the pages of the *Physical Review* swelled with articles repressed during the war, and new journals, like *Nucleonics*, provided an outlet for an outpouring of information. *Nucleonics* was the outgrowth of plans initiated in 1945 with publication of several issues each of three slim mimeographed "magazines," *Atomic Power*, *Atomic Engineering*, and *Nucleonics*. McGraw-Hill attempted to publish a periodical called *Atomic Power* in 1946, but apparently it was premature and ceased publication after three issues.

After the army proposed to continue military control of nuclear research, rank-and-file nuclear scientists lobbied Congress to create a civilian authority instead (Smith, 1971). Their efforts led to establishment of the Atomic Energy Commission (AEC) in 1946. The commission took the reins of one of the largest industrial and engineering complexes in the world. The AEC inherited the MED security and classification system, and a number of enhancements by Congress, which undertook to embargo export of all nuclear information, despite wartime agreements with the British for postwar cooperation. Congress also classified all information developed in working with nuclear fission and the fissile elements, until it could be reviewed (Hewlett, 1981). This congressionally mandated extension of the cloak of secrecy automatically classified as restricted data even information developed outside the secret scientific community. It required that creators of restricted data acquire security clearances if they were to continue to have access to it (Green, 1981; Groves to Tolman, 1945; Tolman to Groves, 1946).

H. Manley, the first secretary of the AEC's General Advisory Committee, recognized the expansion of the realm of classification in 1949. In a manuscript intended for publication in the *Bulletin of the Atomic Scientists*, he wrote:

> Science . . . especially portions of biology, chemistry, mathematics, medicine, metallurgy, and physics, is developing in this country and also abroad along two paths, restricted and open, classified and unclassified. The situation in which scientific work was, in general, freely published no longer exists and at least three nations have laws which restrict the freedom of interchange of certain types of scientific information. Undoubtedy [sic] in terms of numbers of scientific workers affected . . . the United States stands foremost." (1950, p. 1)

Congress, in its extension of secrecy in science, went far beyond Groves's efforts. It also created a precedent for other Cold War efforts to protect America's technological superiority through classification and compartmentalization, which had traditionally been restricted to wartime situations. Long and Roland (1994) surveyed the early history of secrecy and found little evidence for its use in peacetime before the nineteenth century. So strict was this imposition that when the Atomic Energy Act was revised in 1954, the Department of Defense lobbied for loosening it. Relations between the AEC and DoD with respect to nuclear weapons were unsettled after the Soviet atomic bomb explosion in August 1949, and the DoD sought a greater voice in nuclear weapons policy. Since its personnel did not have access to "Restricted Data," the DoD unsuccessfully sought to remove this classification, although it did gain access to "Formerly Restricted Data" (Maus, 1996).

New custodians had to be charged with safeguarding restricted data from the military and other unauthorized parties. The AEC's Technical Information Division (TID), created in the fall of 1947, enjoyed the luxury of a plant located in Oak Ridge for printing classified and unclassified technical reports and the burden of a declassification branch that supervised the activities of the scientists and engineers responsible for advising on declassification. "Senior responsible reviewers" included W. D. Johnson for the plutonium project, Robert L. Thornton for electromagnetic separation, Walter F. Libby for the diffusion process, Manley for weapons, and Harold A. Fidler as secretary. "Standing ad-hoc subcommittees" on chemistry and metallurgy, theoretical nuclear physics, and reactors were responsible to assist the committee, which met eleven times between July 1946, when Groves appointed it, and June 1949. Fidler later became the AEC's Chief of Declassification (Manley, 1950).

Despite their efforts, at the beginning of 1948, the first chairman of the AEC, David Lilienthal, felt the need "to get us in a position where we will really do something about this secrecy incubus. Now when we are being criticized . . . for keeping secrets . . . we are in a position for the first time to . . . junk a lot of this monkey-business" (Lilienthal, 1964, p. 442; U.S. Atomic Energy Commission, 1947–1948). Lilienthal felt that secrecy was abused by those in the military and elsewhere who used it to prevent honest debate on atomic energy issues.

The first test of a Soviet nuclear weapon in August 1949, suggested that the secret design of the atomic bomb had been stolen, despite efforts to prevent the transfer of vital defense information. A secret debate about whether to pursue development of the hydrogen bomb ensued, pitting Oppenheimer and the AEC against Ernest O. Lawrence, Edward Teller, and the defense establishment. In January 1950 Klaus Fuchs confessed his extensive espionage at Los Alamos during and after World War II, helping to resolve the debate in favor of the advocates of a "super" bomb (Williams, 1987; Moss, 1987). After the collapse of the Soviet Union the existence of another physicist-spy was disclosed by the KGB, which identified him only as "Perseus," and who was subsequently disclosed to be Ted Hall by Joseph Albright and Marcia Kunstel (1997). The classic account of the debate is by Herbert York (1976) (see also Bernstein, 1988; Bernstein & Galison, 1989; Hershberg, 1988).

At the same time Executive Order 10104, issued by President Truman on 1 February 1950, officially added the classification level of "top secret" to the existing three levels of secret, confidential, and restricted, and placed the classification system under presidential, rather than congressional, discretion. Congress reacted in a number of ways to the threat, ranging from the witch-hunting crusade of Joseph McCarthy to the requirement of a security clearance for AEC fellowship holders, even when they never used restricted data (Schrenker, 1998; Reeves, 1997; Rovere, 1996).

Spoils of War

The army and the navy, meanwhile, attempted to accommodate vast numbers of German documents captured in 1945. The Department of Commerce, as well as the armed services, participated in the plunder of people and documents well into 1947. Jackson (1992) describes documentation activities by the Air Documents Division, a precursor of the Defense Technical Information Center, involving captured German technical reports related to aeronautical science and technology following World War II (Cf. Lasby, 1971; Bower, 1987; Hunt, 1991; Gimbel, 1990a and 1990b). (For an insightful and comprehensive analysis of U.S. science policy in postwar Germany, see Cassidy, 1994 & 1996. For British activities, see Agar & Blamer, 1998, pp. 224–225.)

Thomas Pynchon's *Gravity's Rainbow* dramatizes the desperate competition between Allied intelligence agencies in occupied Europe at the end of World War II: ". . . the Faithful: the scavengers now following industriously the fallback routes of A4 batteries from the Hook of Holland all across Lower Saxony. Pilgrims along the roads of miracle, every bit and piece a sacred relic, every scrap of manual of verse of scripture" (1995, p. 391). But if Mark Twain in *The Innocents Abroad* found several tons of the True Cross, these visitors found hundreds of times that in "scripture." The documentation gathered in Germany at the end of World War II overwhelmed Allied information and intelligence services.

The information-gathering effort was initiated in 1944 when the Allied Combined Chiefs of Staff ordered a search for war secrets in occupied German territory. Many groups were involved, including several air technical intelligence teams from the navy and army air force. The head of the army air force effort, Caltech aeronautical engineer Theodore von Kàrmàn, gained support from Army Air Force Commander H. H. Arnold for a highly secret project to screen, organize, and catalog 186 tons of documents. The recovered documents were collected in a six-story building at 59 Weymouth Street in London. This "index project" was supervised by twenty-five prominent American scientists and aeronautical engineers under the auspices of the U.S. Navy, U.S. Army Air Force, and British Air Ministry (von Kàrmàn, 1967; Goldman, 1950; Jackson, 1949.)

The personnel of the Air Documents Research Office (ADRC) separated technical from nontechnical documents, sorted technical documents according to source libraries, and constructed "possibly the most rapidly compiled subject heading list in existence" (Jackson, 1949, p. 779). The catalog cards created in this preliminary processing and microfilms of the documents were sent to two hundred agencies. As many as 650 documents were processed daily, and over four tons of documents were screened by the ADRC staff. The ADRC was subsequently transferred to Wright-Patterson Air Force Base in Dayton, Ohio, along with 800,000 documents, and reconstituted as the Air Documents Division (ADD), Air Material Command.

> A more fortunate situation now as to space, personnel, and equipment enabled ADD more closely to approach the ideal "industrial pipeline" make-up. . . . all jobs connected with document processing were analyzed into their elements and lesser skilled persons would be utilized to perform those elements. (One group just established the author entry, another group merely the imprint, another the collation, another the subject headings, etc.) Professional librarians were hired for the document processing procedure to oversee establishment of new subject headings. (Jackson, 1949, p. 780)

Over 55,000 technical reports were eventually processed and combined with the resources of the technical library at Wright-Patterson Air Force Base to form one of the streams that fed a river of military technical reports in the late 1940s and early 1950s. In 1949 the Department of Defense chartered this organization as the Central Air Documents Office (CADO). It was to receive, organize, and distribute those documents of interest to aviation for all three services and to industrial, educational, and research institutions participating in federal aeronautical research and development programs (Goldman, 1950; Jackson, 1949; Jackson 1992).

Captured documents relating to the German atomic bomb project showed only that they had accomplished little, as interrogations and covert recordings of German nuclear scientists confirmed (Goudsmit, 1947; Operation Epsilon, 1993). It was, therefore, more plausible to American politicians that the atomic bomb might remain an American monopoly, and, as Gregg Herken (1980) has shown, they placed their diplomacy and national security on that foundation. Although Dean Acheson, then Undersecretary of State; David Lilienthal, Truman's choice to head the Atomic Energy Commission; and Oppenheimer tried to open up channels for international cooperation with the Soviet Union, they were unsuccessful in averting the nuclear arms race.

Lewis Strauss, one of the first AEC commissioners, "did not share the prevailing state of euphoria as to Stalin's amicable intentions." Strauss sought to resist such pressures for scientific openness, which were based on the Atomic Energy Act's call for "the dissemination of scientific and technical information relating to atomic energy . . . to provide that free interchange of ideas and criticism which is essential to scientific progress." Instead, he supported his position on the basis of the act's prohibition of "exchange of information with other nations with respect to the use of atomic energy for industrial purposes." This stance divided him from the other commissioners: "As I adhered to the letter of the law, the brand of 'security obsession' was early burned upon me, and I still wear it" (Strauss, 1962, p. 256).

Strauss's involvement with the navy's wartime program to develop the proximity fuse convinced him that such weapons could be developed in secrecy. His connection with the Naval Technical Mission persuaded him that the Germans had done so, as well. It

> turned up an astonishingly large and heterogeneous variety of scientific information, material, and people[,] . . . located cunningly concealed laboratories and manufacturing installations (by the ingeniously simple expedient of tracking power lines); [and] found refugee scientists hidden in mines and caves, camouflaged wind tunnels, and rocket plants. It took possession of tons of documents and reports. (Strauss, 1962, p. 149)

Strauss recognized the advantages of a technological lead in nuclear weaponry, and he sought to "preserve that advantage as long as possible by locking up information on atomic energy" (Pfau, 1994, p. 97). He even sought to restrict foreign distribution of radioisotopes, which Strauss believed contained information that might be of use in producing weapons. Oppenheimer ridiculed Strauss's position on radioisotopes during the 1949 "incredible mismanagement" hearings of the Joint Committee on Atomic Energy. He compared their importance to shovels and beer in the creation of atomic energy and ranked them somewhere between electronic devices and vitamins (Pfau, 1994, pp. 108–109). The hearings resulted from revelations that a Communist had received an AEC fellowship, which ballooned into a full-scale investigation of the commission. As a result AEC fellowship holders were required to take a loyalty oath and sign an affidavit that they were not Communists (Marks, 1949; see also Joint Committee on Atomic Energy, 1949).

Strauss launched a campaign within the government to build a super bomb in the wake of the Soviet detonation of their first atomic bomb in the summer of 1949. Aided by Edward Teller and Ernest Lawrence, the Department of Defense, and the Joint Chiefs of Staff, he prevailed over his fellow commissioners and the General Advisory Committee of the AEC in the secret debate over the H-bomb.

Strauss was able to forge a formidable security apparatus within the AEC, of which he became chairman in 1953. The elaboration of the system of classification was accompanied by a tightening of the personnel security system. Everyone involved with atomic energy was subjected to greater scrutiny, and most notoriously in the case of Oppenheimer, many people were deprived of their security clearances. As Eisenhower's choice for chairman of the AEC, Strauss presided over the inquisition of his old enemy. Many other scientists fell prey to the security apparatus of the military and the AEC and lost jobs in industry, academia, and federal laboratories (Pike, 1947; see also Martin, 1946; Engel, 1948; Miller & Brown, 1948; Davies, 1948; Committee on Secrecy and Clearance, 1948; "AEC Criteria," 1947).

Access and Security

Custodians of classified documents had to devise new systems to make scientific and technical information available to those who had a legitimate and legal need to use them. The problems of handling large amounts of classified and unclassified information led CADO to convene a conference in 1949 on the problems of centralized documentation at Wright Patterson. At this time the Air Technical Index, which had been set up in 1947, provided for an automatic, selective exchange of classified information through the Standard Aeronautical Indexing System. It was devised under contract by the Institute of Aeronautical Sciences, which consulted three thousand of its members and two thousand users of CADO in formulating 48 categories and 385 subcategories of technical information, in order to provide guidance to over 15,000 subject headings.

Eugene Jackson of CADO noted that the difficulty of finding information was complicated by military security, which he believed had been developed to protect tactical, strategic, and diplomatic messages. "However," he remarked, "it is coming to the attention of personnel concerned with the military documentation program that scientific materials are being unduly shackled by the imposition of classifications intended for another kind of material" (1950, p. 4). In particular, the dissemination of information was blocked by reserving the authority to declassify documents to the originating agency or individual, both ephemeral in the course of time, rather than allotting it to others conversant with the state of the field:

> CADO has literally hundreds of documents in its collection that it believes are over-classified but which cannot be downgraded now because that agency which prepared the report is no longer in existence . . . the existing military classification directives impose a tremendous obstacle to . . . disseminating technical information." (Jackson, 1950, p. 9)

The Special Committee on Technical Information of the DoD Research and Development Board concurred that technical information should be disseminated promptly to every one, assimilated and correlated with similar material, and made available to all who needed it in their work. Moreover, the committee held that research and development outside DoD needed to be integrated with this database. To avoid duplication of effort, it recommended that "a significant portion of money being spent on research and development be allocated to the specific purpose of creating better methods of insuring that information is recorded and is organized in such a way as to be readily available" (Jackson, 1950, p. 10).

Faced with approximately 4,000 cubic feet of reports, CADO sought to save space through miniaturization. As one military overseer remarked, "Many persons have looked hopefully to the future when all documentation will be done by electronic or other revolutionary methods," but microfilm was still the most convenient means. Not only did it reduce the volume of reports by a factor of ten or more, it also made distribution of multiple copies simpler. Most researchers found this format unobjectionable, and some contractors made full-sized copies from microfilm for internal distribution. The AEC, on the other hand, distributed its reports by printing them after establishing a minimum level for automatic distribution. "The AEC is controlled by the needs of the users, and . . . the user does not desire micro-reproduction" (Warheit, 1950, p. 31). Microfiche was not yet a feasible replacement.

Both CADO and the AEC agreed that IBM

punched-card equipment was a promising tool for cataloging and retrieval of reports. The AEC made particular use of such machines to process classified documents requiring hand receipts. IBM reported that it was investigating the major problems in centralized indexing and searching.

Although CADO was capable of handling approximately 70,000 documents a year, estimates of the total number of reports of interest to military researchers—370,000—drove the DoD to standardize its report formats through interservice agreements, style manuals, and contractual language.

Growth of the AEC Secret Scientific Community

The growth of the AEC following the decision by President Truman to accelerate development of the hydrogen bomb vastly increased the realm of secret scientific communities within its laboratories and production facilities. Livermore was founded as a branch of the University of California Radiation Laboratory at Berkeley in order to serve as a second weapons design laboratory. Production reactors were built at Savannah River to produce more fissile materials. Components of nuclear weapons were produced at Monsanto, General Electric, Pantex, and other new industrial laboratories (Anders, 1987). Although still small compared with the DoD—as were all other government agencies—the AEC was large compared with almost all other public or private enterprises in the United States.

The AEC's security system was very costly. In 1953 the University of California Radiation Laboratory reported security operating costs of $509,079, while Los Alamos Scientific Laboratory spent $383,000 (Reynolds, 1953; Hoyt, 1953). The total did not include "Inefficient Labor Cost while Awaiting Security Clearance," document handling ($42,600), classification and declassification effort ($10,100), overhead, special procurement for security purposes ($23,000), or depreciation of security equipment ($10,000), which brought the total cost to $963,479, a figure Reynolds anticipated would increase as Livermore grew. The total estimated costs for security classification of both federal agencies and federal contractors—which includes personnel, information, and physical security as well as training and management cost—totaled $5.26 billion in 1996. The cost estimates for the CIA were not included because they are classified. Of the agencies reporting, the two accounting for the largest amounts were the Department of Defense at $2.4 billion and the Department of Energy at $92.7 million. This $5.2 billion estimate includes only direct costs and does not include the loss—which many conclude is enormous—that is incurred by the government because of the lack of adequate oversight and open debate of programs that are classified (Garfinkel, 1996; see also Powers,1999).

Many scientists refused to take positions in the laboratories, and a number of scientists within them were discharged. Laboratory contractors complained that the commission reported derogatory information on others who did apply for jobs, "without making a definite recommendation. I judge that the [University of] Chicago's practice is that as soon as the Commission says some one is undesirable, they simply drop them off the payroll and are not inclined to fight back with the Commission as much as Brookhaven has done" (Knox, 1949).

The waning of the "Red Scare," the end of fighting in Korea, and the advent of a president with the prestige and military credibility sufficient to make hard decisions about nuclear weapons relieved the pressure. Over thirty thousand classified documents were in the AEC system when the 1954 Atomic Energy Act and Eisenhower's Atoms for Peace program dictated a revision of the AEC classification guide to make information available for industrial development of nuclear energy. The result was the declassification of eleven thousand and downgrading of eight thousand documents in 1956 and an additional nine thousand in 1957. The AEC also provided 1,404 access permits clearing 22,352 individuals for access to classified documents in order to build nuclear reactors, use isotopes, and mine uranium (Atomic Energy Commission, 1955).

The act also resulted in a stampede to the private sector by entrepreneurial scientists like Frederic de Hoffmann, an important participant in the development of the hydrogen bomb, who left federal service to create the General Atomics Division of General Dynamics (Seidel, 1995). This diaspora exacerbated the problem of classification for the AEC and was, in part, resolved by declassification of subject areas like controlled thermonuclear research, which Strauss made the subject of international display at the Geneva Conference of 1958. Ironically, publication of formerly classified fusion research tempered the interest of industrial concerns in the new technology by revealing how little progress had been made (Bromberg, 1982).

The opening of the closed world of fusion research suggests some of the limits of secret science: among others, lack of peer review, exclusion of politically unacceptable scientists, and lack of international exchanges.

To overcome these problems, secret scientific communities undertook a number of initiatives. Classified meetings, already common at Los Alamos during World War II, became a normal counterpart to the open meetings attended by uncleared scientists. In addition, classified technical journals were established to share information within the secret scientific community of the Atomic Energy Commission weapons laboratories. In this way a simulacrum of the larger world of science was created.

In addition, as the number of classified military technical reports grew, CADO became the Armed Services Technical Information Agency (ASTIA) and undertook to solve the problems of centralized distribution through automation. The result was the Defense Documentation Center (DDC), now known as the Defense Technical Information Center (DTIC), which provided technical report abstracts, work-unit information summaries, research and development planning reports, and independent R&D reports within the closed community of military laboratories and contractors (Defense Documentation Center, 1960).

The history of the development of the Defense Technical Information Center is beyond the scope of this essay (Wallace, 1996; Molholm et al., 1988). To indicate the usefulness of the DOE and DoD secret information systems to historians, however, I reflect below on my own experiences and those of my co-authors using these systems to write the histories of national laboratories and of military laser research and development.

Historian in Classified Worlds

To write a history of the Lawrence Berkeley Laboratory, which included the Livermore branch now known as the Lawrence Livermore National Laboratory, I wanted to see still-classified AEC documents at Livermore, Los Alamos National Laboratory, and DOE headquarters and field offices. The University of California president's office secured an appropriate clearance (of the sort granted to members of the Regents of the University of California whose purview includes both Livermore and Los Alamos). The official historians of the AEC, whose works not only provided a comprehensive guide to the history of the commission but also references to documents used in their research, led me to both classified and unclassified documents important to our study (Hewlett & Anderson, 1962; Hewlett & Duncan, 1962; Hewlett & Holl, 1989). Jack Holl and his staff at the DOE history office assisted us in our studies of Atomic Energy Commission records then held in Germantown, Maryland, and we were also assisted in our research by archivists and technical librarians at Brookhaven National Laboratory, Oak Ridge National Laboratory, and Argonne National Laboratory, as well as those at Berkeley and Livermore. We were also the beneficiaries of the work of Allan Needell and Jane Wolff for the American Institute of Physics, which had produced a number of reports on the DOE laboratory archives (Wolff, 1985; Warnow et al., 1985; Warnow et al., 1982).

The access to classified documents helped fill in many gaps in our understanding of the history of the Lawrence Berkeley Laboratory. The minutes of the Atomic Energy Commission and of its General Advisory Committee provided a national policy context within which we could situate the research and development efforts in the DOE laboratories. However, the materials that were made available in unclassified form deleted much that was still considered secret. Since we had seen the original documents, this was not a problem for our understanding and interpretation of the history, but the admixture of unclassified material with still classified information meant that we could not have had access to the former without the latter. Thus, much of what we learned was not secret but could not have been obtained without a clearance, for, once the text was removed from its classified context, it was often meaningless. Once identified and removed from that context, however, it could be used with an understanding of its significance without revealing information that might harm national security.

The need for access to classified information in writing history related to the AEC and its successors is an artifact of the very broad classification authority given to them by the Atomic Energy Act. This information is "born classified" and, except for historical investigations, is not declassified. The historian cannot investigate it without a clearance, and the efforts of the DOE to declassify large amounts of information, as was done in the 1970s, have been fraught with peril. Wholesale declassification has led to mistakes that have embarrassed, if not compromised, the nuclear weapons community, and consequently, declassifying documents one at a time remains the mode of choice. It remains to be seen if the recent "openness" initiative of the Department of Energy will alter this situation.

I discovered the secret world of DoD information via the Institute for Defense Analysis (IDA) guide, *How to Get It* (Defense Technical Information Center, 1992). I was surprised to find that a whole corpus of scientific literature existed that had not even been mentioned in my courses in the history of science and technology, not to mention the historical literature. When I had an unclassified search of DTIC done by the staff of the University of California Lawrence Berkeley Laboratory, I found what I thought was a bonanza: abstracts of hundreds of technical reports relating to military laser R&D. On a subsequent visit to the Naval Weapons Center (NWC) at China Lake, another outpost of the DoD's secret scientific community, I displayed this treasure and was told that it represented but a fraction of the technical reports available on the subject. The DTIC searcher there, F. Fisher, offered to conduct a more thorough search, which resulted in some 50 cubic feet of material, a large fraction of which was classified. (I had, however, to negotiate with Office of Naval Research [ONR] Security to see these results. It seems that after making the DTIC search, NWC personnel learned I was not a government employee but, rather, a consultant with clearance and left it to ONR to make the decision. The ONR security officer in Washington decided to take the risk without reviewing the material, which was in San Francisco.)

This material served as the basis for my research in the Laser History Project. In the course of that effort I found technical reports in libraries at AEC national laboratories; in air force, army, and navy laboratories; and in the archives of a few defense contractors who granted access to outsiders. A number of military historians offered hospitality. However, when the originating agency had been reorganized or dispersed, and none of its successors was willing to take the responsibility for granting me clearance, I was stymied.

I was able to overcome these difficulties with the aid of the Laser History Project's advisers, who enjoyed sufficient status within the defense community that their intervention and correspondence authorized my access to any document relating to the history of lasers. As one explained to military security officers, "He's a spy, but he's our spy." These letters opened doors from California to Boston.

One Special Access Program report was initially refused despite this comprehensive need to know, until I telephoned one of the authors whom I had recently interviewed and was added to the distribution list. The ONR arranged for a special review of the products of my research to avoid site reviews at every laboratory I visited. The ONR Patent Counsel's Office on Treasure Island provided space for storing and an office for consulting classified reports and interview transcripts. The Patent Counsel, Chuck Currey, was a congenial host for my work for a year and a half. His staff assisted me in the transcription of interviews and made security cabinets available for my work. William Condell of the Laser History Project Advisory Board was responsible for these arrangements, as well as for arranging the initial funding of the project.

The difficulty of transporting the classified tapes of my interviews was largely overcome by ONR's conferral of courier status. (I was told that I must destroy these tapes if my conveyance were hijacked by terrorists, and I debated whether I should eat the tapes or carry a large magnet to degauss them. Both seemed equally conspicuous and potentially lethal remedies, and so I was glad that the occasion did not arise.)

It is obvious that without these efforts on the part of my military patrons to make access and funding available, I would not have been admitted to the secret scientific community of military laser research and development. Once access was granted, however, I also had to assume responsibility for securing classified materials and for making sure that they were reviewed appropriately. This seems to many historians a burden that they should not bear, although scientists and engineers in these communities are accustomed to them. Although tedious and time consuming, these efforts were legitimate and effective, in my view, because of the broad and comprehensive research I undertook.

In addition to technical reports I found many other indications of the creation of secret scientific communities in my research. There are proceedings of classified conferences organized by the Office of Naval Research, the Air Force Office of Scientific Research, and other military sponsors of fundamental research. The weapons laboratories also publish classified journals. Although there are no formal professional societies of secret research to my knowledge, these vehicles perform the same functions for scientists and engineers working in the defense community that meetings and journals do for academic scientists. They permit an exchange of ideas, updates on progress of research, and opportunities for cooperative research.

Viewed from the outside, this community seems to threaten the traditional norms and values of science (see also Foerstel, 1993). From the belly of the beast, however, secret science is merely a different subdiscipline, with its own literature, meetings, laboratories, and concentrations of effort. Journalistic accounts of the community as well as biographies of some of its leaders have appeared, and many scientists who span the boundaries between the worlds of secret and academic science publish in both arenas.

The information systems developed by the Atomic

Energy Commission and the DoD should be seen as part of these communities. Like their patrons, technical libraries partake of both secret and open science and face the problems reconciling the desire to know with the need to know. The formal and informal systems that certify the need to know and the level of classification accessible to a patron are additional procedures that they must observe, but they do not differ substantially from other forms of controlled circulation. For the catalogers, indexers, and abstracters of this information, classified information presents a challenge because their product must be customized for different sets of users, but this, too, is not vastly different from the activities of open information science, as represented by such agencies as the National Technical Information Service (NTIS). Indeed, historian Colin Burke argues that the Cold War transformed the scientific and technical documentation aspirations of the American Documentation Institute into a highly profitable industry (Burke, 1994, p. 211, n. 13).

Profitable, if not perfect. Any number of studies of information systems have pointed out the fragmentation, difficulties, and limits of federal information systems (Committee of DDC Users 1969, pp. 5, 14; Coordination of information, 1961; Auerbach Associates, 1976; "Contract Status Report," 1975). Some of these problems were particularly acute for classified information. One study for the DDC by Auerbach Associates identified the problem in the defense department:

> DDC must be even more concerned than other information transfer organizations (such as NTIS) with the rapid delivery of current information, especially if it is classified. This conclusion is based upon the finding that users with Top Secret classifications found the currency of the information they received less satisfactory than did those with Secret or lower level classifications. DDC is one of the few S&TI [scientific and technical information] and RDT&E [research, development, testing and evaluation] management information services that provide classified information. (1976, p. 34)

The trade-off between speed and security was compounded by the compartmentalization of information such that even the abstracting and indexing services were insular. Indeed, by 1975, one study found that the Advanced Research Projects Agency (ARPA); the Energy Research and Development Administration (ERDA, the successor to the AEC and predecessor of the DOE); the National Aeronautics and Space Administration (NASA); the National Agricultural Laboratory; the National Bureau of Standards (NBS); the National Library of Medicine; the National Oceanic and Atmospheric Association (NOAA); NTIS; the Wright-Patterson Air Force Base Foreign Technology Division; the Air Force System Command's information center; MASIS; the Army Library and Modernized Army Research and Development Information System (MARDIS), and the Naval Material Command's Navy Technical Information (NTII) all provided similar information, used state-of-the-art techniques for information handling, reflected the broad range of information-handling activities of interest to DDC, and processed scientific and technical information, research, development, and testing and evaluation of management information.

Compartmentalization, classification, confusion, and information (C^3I) characterized these systems: "Each major information system has evolved at different points in time to meet different objectives," one study concluded, and added that

> the technologies these systems employ are not readily transferable to other system environments. Previously the diversity among systems resulted in progressive improvements. Today, however, the sheer number of diverse methods of system operations has resulted in difficulties of information exchange [and] in adverse effects among users. (Auerbach, 1975)

Here again, classified information was identified as a culprit in hindering interagency cooperation.

These straws in the wind suggest that the handling of scientific information by many agencies of the secret scientific community continued to frustrate efforts at centralization of scientific information throughout the Cold War. It is perhaps significant that such new agencies as NASA, the Air Force Systems Command, and ARPA all developed their own information agencies, notwithstanding the concern within the Office of the Secretary of Defense to reduce interservice rivalries, duplication of effort, and other dysfunctional aspects of its lack of integration. The continuing bifurcation between DOE and DoD information systems suggests that the centripetal forces of agency growth and differentiation are not easily overcome.

At the dawn of information science T. S. Eliot asked, "Where is the wisdom we have lost in knowledge? Where is the knowledge we have lost in information?" To answer this question literally would seem to be a reasonable objective for the field of science information. The coordination of classification and secrecy with this goal has presented a particularly challenging problem. Secrecy clogs the arteries of our scientific and technical information systems. Radical surgery to relieve this condition in the Soviet Union resulted in the death of the patient. Glasnost has yet to come to the United States.

References

Administration underscores the importance of keeping declassification efforts on track. (1998, November 10). *NCC Washington Update 4*(44), 1.

Agar, J., & Blamer, B. (1998). British scientists and the Cold War: The Defence Research Policy Committee and information networks, 1947–1963. *Historical Studies in the Physical and Biological Sciences, 28*(2), 209–252, especially pp. 224–225.

Albright, J., & Kunstel, M. (1997). *Bombshell: The secret story of America's unknown atomic spy conspiracy.* New York: Harper & Row.

Anders, R. (1987). *Forging the atomic shield: The office diaries of Gordon Dean.* Chapel Hill: University of North Carolina.

Armed Services Technical Information Agency. (1957). Conference on Multiple Aspect Searching for Information Retrieval. Washington, DC: Author.

Article 1a, OSRD Standard Form 1001. (1948). In I. Stewart. *Organizing scientific research for war: The administrative history of the Office of Scientific Research and Development* (Appendix 2). Boston: Little, Brown and Company.

Article 6, OSRD Standard Form 1001. (1948). In I. Stewart. *Organizing scientific research for war: The administrative history of the Office of Scientific Research and Development* (Appendix 2). Boston: Little, Brown and Company.

Atomic Energy Commission. (1947, February). AEC criteria for security clearance. *Bulletin of the Atomic Scientists, 5*(2), 62–63.

Atomic Energy Commission. (1955, January). *Twenty-third semiannual report of the Atomic Energy Commission*. Washington, DC: Author.

Auerbach Associates. (1975, December 12). *Contract status report: Interagency service report*. (AUER-2325/26-TR-2). Charles Babbage Institute Archives, Collection 30. Twin Cities, MN: University of Minnesota.

Auerbach Associates. (1976, June 13*). DDC 10-Year Requirements and Planning Study* (Vol. II). (Contract DSA900-75-C-5161). Charles Babbage Institute Archives, Collection 30. Twin Cities, MN: University of Minnesota.

Balzer, B. (1989). *Soviet science on the edge of reform.* Boulder, CO: Westview.

Baxter, J. P. (1946). *Scientists against time.* Boston: Little, Brown and Company.

Bernstein, B. (1988). Four physicists and the bomb: The early years. *Historical Studies in the Physical and Biological Sciences 18*(2), 263.

Bernstein, B., & Galison, P. (1989). In any light: Scientists and the decision to build the superbomb, 1952–1954. *Historical Studies in the Physical and Biological Sciences 19*(2), 267-347.

Bower, T. (1987). *The paperclip conspiracy: the hunt for the Nazi scientists.* Boston: Little, Brown.

Bromberg, J. (1982). *Fusion: Science, politics, and the invention of a new energy source.* Cambridge, MA: MIT Press.

Burke, C. (1994). *Information and secrecy: Vannevar Bush, ultra, and the other memex.* Metuchen, NJ: Scarecrow.

Burton R. E., & Green, Jr., B. A. (1961, October). Technical reports in physics literature. *Physics Today 14*(10), 35–37.

Bush, V. (1945, July). As we may think. *Atlantic Monthly, 176(1),* 101-108.

Bush, V., & Conant, J. B., to J. R. Oppenheimer. (1943/1983). In David Hawkins. *Manhattan District history, Project Y, the Los Alamos project. Project Y, the Los Alamos story.* Los Angeles: Tomash Publishers.

Cassidy, D. (1994). Controlling German science. I: U.S. and Allied forces in Germany, 1945–1947. *Historical Studies in the Physical and Biological Sciences, 24*(2), 197–236.

Cassidy, D. (1996). Controlling German science. II: Bizonal occupation and the struggle over West German science policy, 1946–1949. *Historical Studies in the Physical and Biological Sciences, 26*(2), 197–240.

Charles, F. (1993). *Operation epsilon: The farm hall transcripts.* Berkeley: University of California Press.

Committee of DDC Users in the Greater Washington, D.C., Area. (1969, September). *Information hang-ups: problems encountered by users of the technical information services offered by DDC and CFSTI, with recommendations for the future*. (Document ED 044 156). Washington, DC: Author.

Committee on Secrecy and Clearance, Federation of American Science. (1948, April 2). Loyalty clearance procedures in research laboratories. *Science, 107,* 333–337.

Coordination of information on current federal research and development projects in the field of electronics: An analysis of agency systems for storage and retrieval of data on ongoing work and of views of private companies on indexing and communication problems, U.S. Senate Committee on Government Operations, 87th Congress 1st Sess. (1961). Washington, DC: Government Printing Office.

Crosland, M. (1962). *Historical studies in the language of chemistry.* Cambridge, MA: Harvard University Press.

Davies, T. H. (1948 July). "Security risk" cases—a vexed question. *Bulletin of the Atomic Scientists, 4*(7), 193–196.

Defense Documentation Center. (1960). *Automation of ASTIA*. Arlington, VA: Author.

Defense Technical Information Center. (1992). *How to get it: A guide to defense-related information resources.* Alexandria, VA: Author.

Eliot, T. S. (1971).Choruses from "The Rock." In *The complete poems and plays, 1909–1950,* (p. 96). New York: Harcourt Brace & World.

Elliott, J. E. (1995, March). Disintegration of the Soviet politico-economic system. *International Journal of Social Economics, 22*(3), 31-60.

Engel, L. (1948, January 17). Fear in our laboratories. *The Nation, 166,* 63–65.

Executive Order 10104. (1949–1953). Definitions of vital military and naval installations and equipment. 15 FR 597, 3 CFR, Comp., 298.

Fagen, M. D., (Ed.). (1978). *A history of engineering and science in the Bell system: National service in war and peace (1925–1975).* New York: Bell Telephone Laboratories.

First report of the Committee on Declassification. (1945, November 17/1949, July 14/rev. 1950, January 6). In J. H. Manley. *Secret science* (Document OA-BEI-71-25) (p. 6). Also enclosed with C. L. Marshall to E. O. Lawrence, (1950, March 29). Documents of Historical Interest from E. O. Lawrence's Files. LBL Archives, LLNL Transmittal. Lawrence Berkeley National Laboratory, Berkeley, CA.

Foerstel, H. (1993). *Secret science: Federal control of American science and technology.* Westport, CT: Praeger.

From glow to flow: A history of laser military R&D. (1987). *Historical Studies in the Physical Sciences 18*(1), 111-148.

Furman, N. S. (1990). *Sandia National Laboratories: The first decade*. Albuquerque: University of New Mexico.

Garfinkel, S., to Rep. D. Skaggs (D-CO). (1996, April 28). The costs of federal classification programs. *The Treasury, Postal Service, and General Government Appropriations Act for 1997. House Report 104-660.*

Gimbel, J. (1990a). *Science, technology, and reparations: Exploitation and plunder in postwar Germany.* Stanford, CA: Stanford University Press.

Gimbel, J. (1990b). The American exploitation of German technical know-how after World War II. *Political Science Quarterly, 105,* 295–309.

Glasnost in Virginia. (1993, October 16). *The Economist*, *329*(7833), A28.

Goldman, L. (1950). Document exchange. In G. R. Hicks (Ed.), *Proceedings of the Conference on Problems of Centralized Documentation* (pp. 52–53). Dayton, OH: CADO.

Goldman, P. (1986). *The end of the world that was: Six lives in the atomic age.* New York: Newsweek.

Goudsmit, S. A. (1947). *Alsos.* New York: Henry Schuman.

Green, H. P. (1981, December). Born classified in the AEC: A legal perspective. *Bulletin of the Atomic Scientists 37*(10), 28.

Groves, L. R. (1983). *Now it can be told.* New York: DeCapo.

Groves, L. R., to R. Tolman. (1945, November 2). Documents of historical interest from EOL's Files. LBL Archives.

Guerlac, H. (1987). *Radar in World War II.* New York: American Institute of Physics.

Heilbron, J. L., & Seidel, R. W. (1989). *Lawrence and his laboratory: Vol. I. A history of the Lawrence Berkeley Laboratory.* Berkeley, CA: University of California.

Herken, G. (1980). *The winning weapon: The atomic bomb in the Cold War, 1949–1950*. New York: Knopf.

Hershberg, J. G. (1988). "Over my dead body": James Bryant Conant and the hydrogen bomb. In E. Mendelsohn et al. (Eds.), *Science, technology, and the military* (pp. 379-430). Dordrecht : Kluwer Academic.

Hewlett, R. (1981, December). Born classified: A historian's view. *Bulletin of the Atomic Scientists 37*(10), 20–26.

Hewlett, R., & Anderson, O. E. (1962). *The new world, 1939/1946* (Vol I). In *A history of the United States Atomic Energy Commission.* State College: Pennsylvania State University Press.

Hewlett, R., & Duncan, F. (1962). *Atomic shield 1947/52* (Vol.II) In *A history of the United States Atomic Energy Commission.* State College: Pennsylvania State University Press.

Hewlett, R., & Holl, J. (1989). *Atoms for peace and war 1953–1961: Eisenhower and the Atomic Energy Commission.* Berkeley: University of California.

Hoddeson, L., et al. (1993). *Critical assembly: A technical history of Los Alamos during the Oppenheimer years, 1943–1945.* New York: Cambridge University Press.

Hounshel, D. & Smith, J. K. (1988). *Science and corporate strategy: Dupont research, 1902–1980.* New York: Cambridge University Press.

Hoyt, H. R., to W. R. Reynolds. (1953, June 17). File "Security—General Correspondence—Atomic Energy Commission." DC Files (Administrative files of the Lawrence Berkeley Laboratory from the files of Donald Cooksey, LBL, and NARA, San Bruno). Berkeley, CA: University of California.

Hunt, L. (1991). *Secret agenda: The United States government, Nazi scientists, and Project Paperclip, 1945 to 1990.* New York: St. Martin's.

Jackson, E. (1949). CADO used varied skills. *Library Journal, 74,* 778–783, 833.

Jackson, E. (1950). Panel discussion, the Air Technical Index. In Hicks, G. R. (Ed.), *Proceedings of the Conference on Problems of Centralized Documentation.* Dayton, OH: CADO.

Jackson, E. B. (1992). Operational information science (documentation) activities at Wright Field, Ohio, before 1950, with emphasis on foreign-language technical reports. *Proceedings of the ASIS Annual Meeting, 29*, 288-90.

Joint Committee on Atomic Energy (1949). *Investigation into the Atomic Energy Commission, report.* Washington, DC: Government Printing Office.

Jones, V. P. (1985). *Manhattan: The army and the atomic bomb.* Washington, DC: U.S. Government Printing Office.

Knox, J. M., to L. J. Haworth, (1949, October 14). BNL Archives, D. O. VII, folder 32. Brookhaven National Laboratory, Upton, NY.

Lanouette, W. (1992). *Genius in the shadows: A biography of Leo Szilard: The man behind the bomb.* New York: Charles Scribner's Sons.

Lasby, C. G. (1971). *Project Paperclip: German scientists and the Cold War.* New York: Atheneum.

Lilienthal, D. E. (1964). *The journals of David E. Lilienthal: The atomic energy years.* New York: Harper & Row.

Long, P. O., & Roland, A. (1994). Military secrecy in antiquity and early medieval Europe: A critical reassessment. *History and Technology 11*, 259-290.

Manley, J. (1950). *Secret science.* (Document 0A-BEI-71-25). Enclosed with C. L. Marshall to E. O. Lawrence, (1950, March 29). Documents of Historical Interest from E. O. Lawrence's Files. LBL Archives, LLNL Transmittal. Lawrence Berkeley National Laboratory, Berkeley, CA.

Manley, J. (1980). A new laboratory is born. In L. Badash, et al. (Eds.), *Reminiscences of Los Alamos, 1943–1945.* (pp. 25–26). Boston: D. Reidel.

Marks, A. W. (1949, June–July) Washington notes: AEC under fire. *Bulletin of the Atomic Scientists, 5*, 163–165, 181.

Martin, L. C. (1946, June). Is secret research justified? *Fortnightly, 159,* 394–400.

Maus, N. C. (1996, July 22). History of classification and declassification. Department of Energy Office of Declassification.

Meigs, M. C. (1982). Managing uncertainty: Vannevar Bush, James B. Conant and the development of the atomic bomb, 1940–1945. *Dissertation Abstracts International* (University Microfilms International No. 82-15951).

Miller, B. S., & Brown, H. S. (1948 February). Loyalty procedures of the A.E.C.—a report and recommendations. *Bulletin of the Atomic Scientists, 4*(2), 45–48.

Molholm, K. N., et al. (1988). The Defense Technical Information Center: Acquiring information and imparting knowledge. *Government Information Quarterly 5*(4), 323-340.

Moss, N. (1987). *Klaus Fuchs: A biography.* New York: St Martin's Press.

Moynihan, D. P. (1998). *Secrecy: The American experience.* New Haven, CT: Yale University Press.

Operation Epsilon: The Farm Hall transcripts. (1993). Introduced by Frank, C. Berkeley: University of California Press.

Pfau, R. (1994). *No sacrifice too great: The life of Lewis L. Strauss.* Charlottesville: University Press of Virginia.

Pike, S. T. (1947, November). Witch-hunting then and now. *Atlantic Monthly, 180,* 93–94.

Powers, T. (1999, February 4). Review of *Secrecy: The American Experience* by Daniel Patrick Moynihan. *New York Review of Books XLVI*:(2), 20-24.

President critical of restrictive declassification provision in defense authorization act. (1998, October 26). *NCC Washington Update 4*(42), 1.

Price, D. J. De S. (1963). *Little science, big science.* New York: Columbia University Press.

Pynchon, T. (1995). *Gravity's rainbow.* New York: Penguin. Reprinted from New York: Viking, 1973.

Reeves, T. C. (1997). *The life and times of Joe McCarthy: A biography* (Reprint edition). Lanham, MD: Madison Books.

Reich, L. K. (1985). *The making of industrial research, science and business at GE and Bell.* New York: Cambridge University Press.

Reynolds, W. R., to J. J. Flaherty. (1953, June 4). Re: survey of security operating costs. DC Files (Administrative files of the Lawrence Berkeley Laboratory from the files of Donald Cooksey, LBL, and NARA, San Bruno). Berkeley, CA: University of California.

Rovere, R. H. (1996). *Senator Joe McCarthy* (Reprint edition). Berkeley: University of California Press.

Schrenker, E. (1998). *Many are the crimes: McCarthyism in America*. Boston: Little, Brown and Company.

Seidel, R. W. (1995, May). Technology transfer: Halfway houses. LANL Report 12927-MS (May 1995). Los Alamos, NM: Los Alamos National Laboratory.

Serber, C. (1988). Labor pains. In J. Wilson & C. Serber. *Standing by and making do,* (p. 65). Los Alamos, NM: Los Alamos Historical Society.

Serber, R. (1993). *The Los Alamos primer.* Berkeley: University of California Press.

Serber, R. (1998). *Peace and war, reminiscences of life on the frontiers of science.* New York: Columbia University Press.

Shannon, C. E. (1993). Communication theory of secrecy systems. In N. J. A. Sloane & A. Wyner (Eds.), *Collected papers.* New York: IEEE.

Shils, E. (1956). *The torment of secrecy: The background and consequences of American security policies.* Glencoe, IL: Free Press.

Smith, A. K. (1971). *A peril and a hope: The scientists' movement in America 1945–47.* Cambridge, MA: MIT Press.

Stewart, I. (1948). *Organizing scientific research for war: The administrative history of the Office of Scientific Research and Development.* Boston: Little, Brown and Company.

Strauss, L. (1962). *Men and decisions.* Garden City, NY: Doubleday.

Tolman, R., to L. R. Groves, (1946, February 18). Documents of historical interest from EOL's Files. LBL Archives.

U.S. Atomic Energy Commission. (1947, July 1–1948, June 30). *Abstracts of declassified documents.* Oak Ridge, TN: Technical Information Division, Oak Ridge Directed Operations.

U.S. Atomic Energy Commission. (1947, July 1–1948, June 30). *Abstracts of declassified documents* (Vols. 1–2). Oak Ridge, TN: Technical Information Division, Oak Ridge Directed Operations.

Update on the declassification amendment to prevent disclosure of nuclear weapons information. (1998, September 30). *NCC Washington Update 4*(38), 3.

von Kàrmàn, T. (1967). *The Wind and beyond.* Boston: Little, Brown.

Wallace, L. E. (1996). The story of the Defense Technical Information Center 1945–1995. Washington, DC: Naval Research Library Technical Information Division.

Warheit, I. A. (1950). Panel discussion on miniaturization in documentation. In Hicks, G. R. (Ed.), *Proceedings of the Conference on Problems of Centralized Documentation* (pp. 23–31). Dayton, OH: CADO.

Warnow, J., et al. (1982). *A study of preservation of documents at Department of Energy laboratories.* New York: American Institute of Physics.

Warnow, J., et al.(1985). *Guidelines for records appraisal at major research facilities.* New York: American Institute of Physics.

Weart, S. (1979). *Scientists in power.* Cambridge, MA: Harvard University Press.

Williams, R. C. (1987). *Klaus Fuchs: Atom spy.* Cambridge, MA: Harvard University Press.

Wilson, R. Accelerators, past and present. (1993). In *Los Alamos National Laboratory, 50th Anniversary Seminar Series* (LANL Technical Report 93-41VT). Los Alamos, NM: Los Alamos National Laboratory.

Wolff, J. (1985). *Files maintenance and records disposition: A handbook for secretaries at Department of Energy laboratories.* New York: American Institute of Physics.

Wooster, H. (1987). Historical note: Shining palaces, shifting sands: National information system. *Journal of the American Society for Information Science, 38*(5), 321–335.

York, H. (1976). *The advisors, Oppenheimer, Teller, and the super bomb.* New York: W. H. Freeman.

Zachary, G. P. (1997). *Endless frontier: Vannevar Bush, engineer of the American century.* New York: Free Press.

Science and Scientific Information Systems

The Role of the Scientific Paper in Science Information Systems

Bernd Frohmann

Abstract

Studies of scholarly communication among scientists agree that the journal article is peripheral to research-front science. This conclusion derives from a cognitivist, science-as-knowledge model that construes scientific work as a conceptual activity in which information plays a central role. But according to recent studies of scientific practices, scientific work is not primarily conceptual but instead consists in stabilizing complex networks of heterogenous elements. These studies suggest that discursive elements and practices, such as writing and using journal articles, contribute to the stability of networks. Thus the importance of the article to research-front science need not consist in its role in the communication of information. An especially important analytical category of studies of scientific practices is that of an *objectifying resource*. This paper argues that the journal article is central to research-front science because it is among science's objectifying resources. The argument proceeds by exploring the implications for science information systems of three historical analyses of science: Sir Francis Bacon's model of state-organized science; Robert K. Merton's analysis of journals as systems of credit and reward; and Robert Boyle's literary technology for warranting scientific facts. Some conclusions of recent studies of the relationships between contemporary laboratory practices and the discursive practices resulting in the production of the journal article strengthen the argument.

Introduction

Ever since J. D. Bernal's controversial proposal to the Royal Society's Scientific Information Conference of 1948 for central distribution of scientific papers (Royal Society, 1948), the scientific journal article has occupied a precarious position in studies of science information systems. Journals have been called "the most important source and medium of scientific information" (Mikhailov, Chernyi, & Giliarevski, 1984, p. 198). Indeed, "scientific documents are a form of science. Without them, science cannot exist" (p. 147). Yet the exponential growth of scientific documents, rather than signifying a corresponding growth in scientific communication, signals to some the collapse and ruin of the entire system. In a typical assessment, written almost a decade before his Royal Society study, Bernal lamented of "the chaos of scientific publication" that "the burden of this vast mass of [journal publication] is in itself a great handicap to scientific research" (1939, pp. 118, 119). Even worse, the significance of the journal article has been challenged apart from its rapid proliferation. Restating a commonplace observation, Robert Hayes concluded that "natural scientists are focused on the acquisition of new data rather than the analysis of existing records. For them, the records of the past are peripheral to research" (1992, p. 6). Price's index (Price, 1970; reprinted in Price, 1986) gave mathematical expression to a recurrent conclusion of studies of science information systems that the archive of "the records of the past" is consulted with such a low frequency that other, and most typically, informal channels of scientific communication are regarded as central to research-front work.

This paper argues that the paradox of a document form both essential and marginal to science arises from a cognitivist conception of science, which Andrew Pickering calls the science-as-knowledge model. It interprets science as a unified conceptual field whose structure is determined by scientific method:

> For the logical empiricist, say, scientific culture consists in a field of knowledge and knowledge claims, and scientific practice consists in the appraisal of conceptual knowledge claims against observational knowledge, an appraisal ideally governed by some logic or method. (1992, p. 3, note 2)

In such a model "scientists figure as disembodied intellects making knowledge in a field of facts and observations" (Pickering, 1995, p. 6). It has been criticized as an "algorithmical model" of knowledge, one "very much in accord with the view of the information scientist," who "views knowledge as the sort of information that enables a computer to carry out its programmer's intentions" (Collins, 1992, p. 75). When knowledge is seen as "a set of formal instructions, or pieces of 'information' " (p. 57), the salient activity of science becomes *information processing*. Scientific method becomes the program that generates propositions from scientific information. Insofar as the production of scientific knowledge is communal and cumulative, information must be communicated among scientists. Documents enter the picture as vehicles for the communication of information, which is interpreted as the epistemic content of the documents' statements. The problem for the study of science information systems is to analyze traffic flow in epistemic content, or "information," in order to maximize knowledge production.

Given this analytical framework, the role of the journal article becomes paradoxical. One response to the paradox is to accept the science-as-knowledge model, but to locate the article's importance in a social system of credit and reward. This response is problematic because it uncouples the labor of scientific writing from laboratory work and knowledge production. Another response is to reject the model, locating the article at the center of scientific labor, but at the expense of the centrality of information flow to scientific work. This paper argues, by three historical examples, for the second response. first, the historical antecedents of the science-as-knowledge view are briefly indicated in Sir Francis Bacon's model of the communal scientific enterprise. Second, the mid-twentieth century response in terms of credit and reward is located in the work of Robert K. Merton. finally, it is argued that the concept of an objectifying resource, which is central to contemporary studies of scientific practices, suggests a more plausible analysis of the journal article's role in scientific work. The historical roots of this concept reach back to Boyle's contribution to the development of a literary technology that helps transform local laboratory results into phenomena of a shared, objective world.

Documents and Information in Baconian Science

Bacon's model of science, emerging from the fragments of his projected *Instauratio magna*, the "great instauration," or renewal, of the sciences in the early seventeenth century, anticipates some important aspects of modern notions of science information systems. Although the origin of the scientific journal is usually dated some forty years after Bacon's death in 1626, he anticipated science's literary technology by placing a system of written records at the heart of knowledge production.

In his "Plan of the Great Instauration" Bacon (1960, p. 17) insists that its first part must be a "Division of the Sciences." He describes this as "a summary or general description of the knowledge which the human race at present possesses" (1960, p. 17). In other words the Division of the Sciences is to be a written record of what is currently known. Its most fundamental division reflects "the absolute chasm which exists between the truths given in revelation through the Word of God and axioms discovered by the powers of man and, secondly, through distinctions among the human faculties" (Anderson, 1948, pp. 148–149).[1] The main classes corresponding to the three main human faculties—memory, imagination, and reason—are history, poetry, and philosophy. History is divided into natural and civil. Natural history records and organizes the phenomena of nature. It is a classification of the epistemic content of records, which, once written out according to strict rules designed to purge them of anything other than what may be derived through observation and experiment, provides the basis for inductive generalizations. In Anderson's gloss natural history is a "delineation of the sort of experimental history which is suitable for the building of a philosophy" (Anderson, 1948, p. 259). Because "it provides the materials on which the understanding is to operate" (p. 260), this part of the classification is so necessary "to the Instauration that, if it cannot be provided, the scheme cannot become operative and the whole project for the reform of knowledge may as well be given up" (p. 259).

In a statement clearly expressing his science-as-knowledge approach, Bacon says of his Division of the Sciences: "However, I take into account not only things

[1]The details of Bacon's classification must be pieced together from several writings, since the Division of the Sciences was never completed. Bacon offered in its place a Latin, reworked version of his much earlier *Advancement of Learning* (1605), titled *De dignitate et augmentis scientiarum*. For an account of the texts and the details of Bacon's classification, see chapters 13 and 14 of Anderson (1948).

already invented and known, but likewise things omitted which ought to be there" (1960, p. 18). His classification can include things "which ought to be there" because its principles do not derive from literary or cultural warrant, but from the structure of knowledge itself. Its logical foundation means that it is hospitable to the inferences drawn from existing records according to Bacon's proposals for reasoning correctly and generating higher-order conclusions from "things already invented and known." Since the progress of thought is from natural history to natural philosophy, the classification's hospitality to the things "which ought to be there" creates a class for records in the third part of the Great Instauration, "The Phenomena of the Universe, or a Natural and Experimental History of the Foundation of Philosophy."

The relationship between Bacon's classification and its literature inverts that of nineteenth- and twentieth-century systems.[2] In Bacon's system a classificatory position does not derive its warrant from the literature but from the organization of the natural world as represented by scientific knowledge. Because new information is generated from previous information by scientific reason, new scientific records have a position already guaranteed for them in the classification of documents, corresponding to the position of their epistemic content in the organization of knowledge.

In *New Atlantis*, Bacon's utopian fable of state-organized natural science, the knowledge of the natural world is produced only through a highly structured social system. The "things already invented and known" are collected for inclusion in the Division of the Sciences principally from books and other written records. This document collection activity is divided among several different ranks of scientific worker: "Merchants of Light" (those who sail to distant lands to collect and make reports of experiments), "Depredators" (those collecting local experimental reports), and "Mystery-Men" (those collecting reports of the mechanical arts). The information gathered and recorded by workers in these first three divisions of scientific labor is then processed by those in the three following divisions. The "Compilers" re-present previous experiments in "titles and tables," displaying observations perspicuously, thus allowing axioms to be more easily drawn from them. Reasoning from the records of the Compilers, and consulting with other scientific colleagues, the "Lamps" suggest new experiments to advance knowledge by building upon current and previous work. The experiments are performed by the "Inoculators," whose reports are then submitted to the "Interpreters of Nature." The Interpreters generate higher-order axioms to guide further observations. Finally, the "Dowry-Men," or "Benefactors," apply the knowledge gained to useful inventions.

Bacon's description of the scientific enterprise as a set of collaborative, socially organized activities of gathering, producing, processing, classifying, and applying written records constitutes what we would today call a science information system. He recognized that science does not develop merely by thought, experiment, and observation, but requires a literature.[3] For him, a scientific record contributes a unit of scientific knowledge. Knowledge does not advance merely by an increment in the number of its constituent units, but by the organization of the units through the inferences and generalities—axioms—drawn from them, such that new observations can be made and further experiments devised. The proper classification of recorded units of knowledge is not merely heuristic, allowing higher-order axioms to be drawn from them; it is also representational. The differentia of the subclasses of natural history and natural philosophy are categories that represent the structure of the natural world. Categories of this kind achieve the goal of the classification, which is to facilitate generalizations that increase our knowledge of nature.[4]

What would the structure of scientific literature look like, given Bacon's view of the scientific enterprise? Journal articles would be organized by a classification system reflecting the levels of generality of the axioms derived from the experimentally generated observations reported in them. The structure is hierarchical, culminating in a set of articles containing high-level generalizations. The organization of documents mirrors the structure of knowledge, because the imperatives of document classification derive from the inductive inferences holding between classes of recorded information.

Bacon's model of scientific activity resists many of the reductivist tendencies of science-as-knowledge models. For Bacon, science is the product of much more than merely cognitive activities. He emphasizes the social organization of collective labor, strict rules for writing

[2]For a brief but illuminating recent discussion of nineteenth- and twentieth-century classifications, see Miksa (1998).

[3]The importance of documents, their collection, organization, and the social structure required for the production of knowledge from them, is emphasized in Martin's study of the relationships between Bacon's view of science and his design for the British imperial state (Martin, 1992).

[4]See Anderson's gloss on the categories of the natural philosophy class (1948, pp. 154–156).

scientific documents, and applications for the production of machines, instruments, and other technological devices. Yet the organizing principle of scientific activity derives from a science-as-knowledge model. Scientific documents communicate information, in the form of observations, that provides raw material for new results. Since observations support inductive generalizations only as members of a class of observations, they must be combined with other observations. Since many of these observations derive from the literature, the epistemic content—we would say, the "information"—conveyed by documents is as directly implicated in the production of new results as the information generated through experimental work. Science is a conceptual field, a systematic organization of information and the propositions derived from them. It consists of immaterial, conceptual entities: information in the form of observations, concepts, and propositions. The essentials of Bacon's model are threefold: 1) Information is identified with the epistemic content of documents; 2) document classification mirrors the classification of information, and both are based upon the organization of knowledge; and 3) the communication of scientific information is achieved by the system of scientific document production, organization, and use. In this model classifiers map and mirror the structure of knowledge. They labor alongside experimental scientists as coworkers in the production of knowledge.

Cognitive Contamination: Merton's Norms

Although Bacon's science information system implicates a complex social organization responsible for the production, organization, and circulation of documents, it is governed by the cognitive imperatives of science-as-knowledge models. But for Merton, the father of the sociology of science, the structure of scientific knowledge is not sufficient to regulate a science information system. He sees science as a social order whose cohesiveness, stability, and systematic advance depend not only on the epistemic value of scientific information but also on shared values based on adherence to specific norms. Since Merton's norms are treated in detail in a voluminous literature,[5] his original four are simply listed here: organized skepticism (scientists are expected to evaluate new knowledge critically and objectively); disinterestedness (their findings are not expected to be used in a self-interested fashion; they are expected to maintain an attitude of emotional neutrality toward their work); universalism (scientific merit should be evaluated independently from the personal or social qualities of the individual scientist); and communalism (since scientists do not own their findings, secrecy is forbidden, and open communication is prescribed).

Since Merton's social norms—what he called "the ethos of science"—build moral imperatives into the heart of scientific activity, knowledge production comes to depend on more than adherence to cognitive and technical standards. The scientist not only follows rigorous methodological precepts, such as those Bacon took great pains to elaborate, but also works at "fashioning his scientific conscience" by cultivating an ethos, "that affectively toned complex of values and norms which is held to be binding on the man of science" (Merton, 1973b, pp. 268–269). However, Merton also recognized that since abiding by the ethos of science is not its own reward, scientists need to be acknowledged for observing the norms. Whereas Bacon's system fails to build rewards into the social structure of science, Merton's system embeds reward in science's information system. The most important kind of reward for scientific work is "eponymy, the practice of affixing the name of the scientist to all or part of what he has found." The most highly prized rewards are therefore titles, such as the Copernican system, Hooke's law, and Halley's comet, but

> The large majority of scientists, like the large majority of artists, writers, doctors, bankers and bookkeepers, have little prospect of great and decisive originality. For most of us artisans of research, getting things into print becomes a symbolic equivalent to making a significant discovery. (Merton, 1973c, p. 316)

Not all publications, however, have equal value, since, as Merton points out, "for a published work to become a genuine contribution to science, it must, of course, be visible enough to be utilized by others" (1973a, p. 332). Yet for the great majority of scientists, mere publication becomes the chief form of eponymous recognition and reward because the mechanism of publication—the referee system—is an "institutionalized pattern of evaluation."[6] Since journal referees bestow or withhold the imprimatur of science, they administer one of science's most important reward systems. Referees are

[5]For a short and accessible introduction to the "scientific ethos debate," see Toren (1983); a useful list of references may also be found in Bazerman (1983, p. 168).

[6]Garvey also emphasizes the reward inherent in mere publication, through the "use of journal articles as the primary source to establish

"an example of status-judges who are charged with evaluating the quality of role-performance in a social system . . . Status judges are integral to any system of social control through their evaluation of role-performance and their allocation of rewards for that performance" (Merton & Zuckerman, 1973, p. 461).

If science's social system is structured by normative standards, then how is compliance with its cognitive and epistemic standards to be guaranteed? Merton's answer is that abiding by the ethos of science advances knowledge because the norms flow from scientific method:

> The institutional goal of science is the extension of certified knowledge. The technical methods employed toward this end provide the relevant definition of knowledge: empirically confirmed and logically consistent statements of regularities (which are, in effect, predictions). *The institutional imperatives (mores) derive from the goal and the methods* [emphasis added]. The entire structure of technical and moral norms implements the final objective. The technical norm of empirical evidence, adequate and reliable, is a prerequisite for sustained true prediction; the technical norm of logical consistency, a prerequisite for systematic and valid prediction. The mores of science possess a methodologic rationale but they are binding, not only because they are procedurally efficient, but because they are believed right and good. They are moral *as well as* technical prescriptions. (Merton, 1973b, p. 270)

If Merton's norms are connected as he claims with science's cognitive and technical imperatives, then insofar as rewards are distributed through science's formal information system, the Mertonian model places great stress on the epistemic value of the scientific journal article. Reward through publication recognizes work of epistemic value only if journal articles reflect scientific work accurately and are used to further scientific knowledge. If they do neither, then the reward system becomes uncoupled from epistemically valuable activity. Merton often expresses his agreement with both points by assuming that formal publications are used directly in knowledge production:[7]

> The system of monitoring scientific work before it enters into the archives of science means that much of the time scientists can build upon the work of others with a degree of warranted confidence. It is in this sense that the structure of authority in science, in which the referee system occupies a central place, provides an institutional basis for the comparative reliability and cumulation of knowledge. (Merton & Zuckerman, 1973, p. 495)

If journal articles have the epistemic value Merton assumes, then his model is compatible with Bacon's. Given the imperfections of mortals, a system for distributing rewards for submitting to the discipline of scientific method is needed. A nice solution is Merton's: Embed the reward system in science's formal information system. With the norms in effect the circulation of journal articles not only communicates the information required for the performance of advanced scientific work, as Bacon's model requires, but also distributes rewards to information of genuine epistemic value, thereby satisfying Merton's model.

The problem, however, is that the epistemic value of a reward system embedded in the formal channels of science's information system is held hostage to the question of whether journal articles contribute information used directly in the derivation of new results. Yet studies of scholarly communication in science show that they only rarely convey the information required for research-front work. The possibility that Merton's ethos of science can become unhinged from the epistemic value of the information conveyed by the communication system in which it is embedded therefore introduces a destabilizing element into his analysis. Rewards through publication institutionalize the norms of science only if publications actually convey the information used in the derivation of new knowledge. Otherwise, the reward system floats free, as it were, from epistemic value. The system may continue to function even if journal articles are used only by referees or status-judges to bestow the reward of publication upon articles whose value as information for deriving new knowledge has dropped to

priority" (1979, p. 75): "In almost every scientific discipline today, the socially accepted medium for establishing priority is the scientific journal article" (p. 69).

[7]The "information-recognition exchange model of scientific organization" developed by Hagstrom (1965), Merton's student, according to which the "organization of science consists of an exchange of social recognition for information" also exhibits the same dependency of the integration of moral and cognitive value on the logical role of the information exchanged for social recognition. When manuscripts are given as "gifts," in Hagstrom's analysis (pp. 12–23), in exchange for the social recognition granted through publication, information of epistemic value to scientific knowledge is thereby rewarded only if the gift *is* information, that is, if it is logically related to the derivation of new knowledge. Hagstrom makes this explicit in his comments on another form of recognition operating through the formal channels of scientific communication, the practice of citing the publications of others: "It is usually necessary, even obligatory, for them to recognize previous work, *for the validity of their own contributions depends logically on the earlier work*" (p. 24; emphasis added).

zero.[8] Thus Garvey, for example, can assert, without violating the Mertonian project, that the information communicated in journal articles is not useful at the research front where new knowledge is generated; yet its mere publication constitutes reward by establishing priority and ownership.[9]

A strength, albeit unintended, of Merton's model is its capacity to explain the importance of documents largely useless in the production of new scientific knowledge.[10] But this explanatory power is gained only at the expense of the communicative and informational value of the overt content of the scientific journal article, which played such a central role for Bacon. The destabilizing element of Merton's analysis relocates the informational value of the article in a latent content, which helps position its author in a social hierarchy of status and prestige.

The Journal Article as an Objectifying Resource

Science-as-knowledge models privilege the role of information because they emphasize such activities as data generation and processing, constructing and testing hypotheses, and theory building. They marginalize the role of the journal article in science information systems because studies of scientific communication show that articles are not the source of the information required for the production of research-front knowledge. Furthermore, content analyses of journal articles show that they do not represent the process of scientific discovery, but present after-the-fact proof, omitting false leads, unsuccessful efforts, and the factors resulting in both the choice of problem and the final set of procedures. Not only do they typically fail to provide enough information for the replication of successful procedures, but also the very possibility of replication has been challenged (Collins, 1992).[11] These problems are nicely condensed in Bazerman's question:

> If a scientific paper is not a complete account of a scientist's observations and doings, nor a tightly argued deductive proof of claims, nor an unproblematic conveyor of claims to be objectively evaluated fairly and promptly by a professional audience, what indeed is the scientific paper communicating, and to whom? (1983, p. 158)[12]

Contemporary studies of scientific practices have rejected science-as-knowledge models. Pickering (1992, p. 6), for example, asks whether "analytic repertoires developed in the service of a problematic of knowledge can serve as the primary basis for understandings of practice." He concludes that "most scholars who have taken it as their task to get to grips with scientific practice in some detail have found that they cannot." Modeling science as a conceptual field "does not offer much purchase upon the complexities evident in the nearest laboratory" (p. 5). Studies of scientific practices emphasize instead the "patchiness," or the "motley" of science rather than conceptual homogeneity or unity: "Scientific culture is made up of all sorts of bits and pieces—material, social, and conceptual—that stand in no necessary relation to one another" (p. 8). The varieties of scientific practices and the complexities of scientific culture bring into sharp relief the false assumption of science as a unified, conceptual field: "Scientific culture is *disunified, multiple, and heterogeneous*" (Pickering, 1995, p. 70). From this view the goal of scientific work is not the pro-

---

[8]In an exchange between Harnad and Fuller on electronic journal publishing, Hanad argues that the "esoteric" literature, i.e., scholarly journal articles, has no market: "Esoteric serial publishers will learn that their real clients are esoteric authors (actually, their institutions and granting agencies) rather than readers" (Harnad, 1995b, p. 311); the "captive audience" of the journals "is not the readership of the journals, it is the institutional library that must have the entire journal in hand for the few, if any, who ever consult any particular article" (p. 317). In his response to Harnad (1995a), Fuller also emphasizes the noncognitive function of the journal literature: "The communication of results, the allocation of credit, and the creation of an archive all reflect the *publicity* function of journals" (Fuller, 1995a, p. 300). For the final word on the exchange, see Fuller (1995b).

[9]Garvey notes that his studies with N. Lin and K. Tomita "raise some questions about the function of current journal articles: Can the journal article any longer be regarded as a vehicle which effectively communicates current scientific information? If not, can the journal article be reworked to function more efficiently in the capacity of integrating scientific information into a larger framework?" (1979, pp. 223–224.) These are the same questions posed by Bernal forty years earlier.

[10]Even this strength may be challenged. If scientists use the results of others before they get into print, then their knowledge of priority and ownership does not depend upon journal publication.

[11]Hacking notes that experiments do not generally replicate previous work in order to refute theoretical conjectures: "Folklore says that experiments must be repeatable . . . roughly speaking, no one ever repeats an experiment" (1983, p. 231).

[12]Bazerman's paper is a useful introduction to some of the pre-1980 literature on the sociology of science and its implications for scientific and technical writing. Bazerman is one of the few who have made the role and function of the scientific paper a distinct research topic (1988). See also Knorr-Cetina (1981); for actor-network approaches to scientific writing, see Callon, Law, and Rip (1986); Latour and Woolgar (1986); and Latour (1987).

duction of a conceptual field but the stability of networks consisting of many heterogeneous elements.

Bazerman (1994, p. 118) has suggested extending the idea of networks, or the "notion of system . . . to include all kinds of symbolic representations, relationships, practices, and objects that must be brought into alliance for any technology or scientific knowledge to take hold." If the notion of system is extended in the way he suggests, then such discursive elements as the journal article belong to the "motley" of the natural sciences. The problem of their role in scientific activity then shifts to their contribution to stabilizing networks. Alternatives to information must be found among the concepts that explicate this role. Chief among them is the concept of an objectifying resource, which suggests that the article aids stability by its contribution to the construction of the objectivity of the natural world.

The historical origin of the article as one of the most important *discursive* objectifying resources of science may be traced to Boyle's contribution to the "literary technology" devised in the seventeenth century for reporting scientific results. According to Steven Shapin and Simon Schaffer (1985, p. 76), it was one of the "three technologies . . . involved in the production and validation of matters of fact: material, literary, and social." Boyle recognized the importance of discipline in the development of the experimental report as a particular literary form. Among its important rhetorical features were "virtual witnessing" and a moral posture of "modesty." To borrow a term from Donna Haraway, "modest witness" is an apt name for this literary style.[13]

Virtual witnessing was the literary equivalent of the careful staging of scientific experiments. Once a pertinent phenomenon was produced in the laboratory, it was reproduced before a highly select group of witnesses. Such demonstrations "were a routine feature of the meetings of the Royal Society, and a *Register-Book* was provided for witnesses to testify their assent to experimental results" (Shapin, 1996, p. 107). Such direct witnessing, although important to the constitution of matters of fact, was a limited way of propagating a new and highly disciplined form of experience that was to legitimate scientific assent.[14] Boyle therefore sought to multiply witnesses through "the production in a *reader's* mind of such an image of an experimental scene as obviates the necessity for either direct witness or replication" (p. 60). In order to achieve such virtual witnessing, a specific literary technology had to be devised, "a technology of trust and assurance that the things had been done and done in the way claimed" (p. 60). Boyle realized that if "one wrote experimental reports in the correct way, the reader could take on trust that these things happened. Further, it would be as if that reader had been present at the proceedings. He would be recruited as a witness and be put in a position where he could *evaluate experimental phenomena as matters of fact*" (p. 63; emphasis added). For Boyle, this took the form of a literary style characterized by an "ornate sentence structure, with appositive clauses piled on top of each other," in order "to convey circumstantial details and to give the impression of verisimilitude" (p. 63). This ornate, rather than succinct, style was required to present simultaneously, in one snapshot, as it were, all the details required for virtual witnessing. "Elaborate sentences, with circumstantial details encompassed within the confines of one grammatical entity, might mimic that immediacy and simultaneity of experience afforded by pictorial representations" (p. 64).

Neither direct nor virtual witnessing were forthright presentations of the highly localized and contingent laboratory circumstances that contemporary studies of scientific practices have revealed as typical elements of scientific work. Boyle's "circumstantial style" was designed as the prose version of the *staged* experimental scene, one already purged of the local context, contingencies, situatedness, and opportunistic reasoning involved in the actual production of the laboratory phenomenon. Thus the experimental report, carefully designed for virtual witnessing, is the discursive correlate of a theatrical strategy of objectivity. Since the point of the literary technology is to substitute for replication and direct witnessing, its circumstantial details must be as routinized and standardized as those of the staged event. The experimental report is written to present an objective phenomenon of the natural world, not a local phenomenon arising from the kind of contingencies

[13]Haraway's use of this expression is ironic, as it serves to articulate her criticism of the gender blindness of Shapin and Schaffer's work (Haraway, 1996).

[14]The extent of this limitation is evident in the following observation: "For practical reasons alone the number of direct witnesses for experimental performances was always limited: in Boyle's laboratory that public probably consisted of at most three to six competent colleagues, and audiences for Royal Society trials rarely exceeded twenty and were typically much smaller" (Shapin, 1996, p. 107).

encountered by modern ethnographers who study laboratory work.[15]

The second important rhetorical feature of Boyle's literary technology is "the modesty of experimental narrative":

> It was the burden of Boyle's literary technology to assure his readers that he was such a man as should be believed. He therefore had to find the means to make visible in the text the tokens of a man of good faith. . . . Thus the literary display of a certain sort of morality was a technique in the making of matters of fact. A man whose narratives could be credited as mirrors of reality was a *modest man*; his reports ought to make that modesty visible. (Shapin, 1996, p. 65)

To strike a posture of modesty through scientific writing consists, first, in eschewing grand, natural philosophical systems in favor of the piecemeal work characteristic of the scientific journeyman satisfied with the limited goals of experimental reports. "Those who wrote entire systems were identified as 'confident' individuals, whose ambition extended beyond what was proper or possible. By contrast, those who wrote experimental essays were 'sober and modest men,' 'diligent and judicious' philosophers, who did not 'assert more than they can prove' " (p. 65). And proof in experimental matters required that all traces of personal style be purged from the writing so that the facts could appear to speak for themselves. Thus a

> technique for showing modesty was Boyle's professedly "naked way of writing." He would eschew a "florid" style; his object was to write "rather in a philosophical than a rhetorical strain." This plain, ascetic, unadorned (yet convoluted) style was identified as *functional*. It served to display, once more, the philosopher's dedication to community service rather than to his personal reputation. (p. 66)

To pursue similarities between contemporary scientific writing and Boyle's literary technology of virtual witnessing is not to suggest that the experimental report has not changed since the seventeenth century. But even today grand schemes, typically published in books, have a lower epistemic status than journal articles. Furthermore, the decontextualized style of the Methods and Materials and the Results and Discussion sections of the contemporary journal article also consists in a flat, unadorned, recitation of events (Knorr-Cetina, 1981). Perhaps the most striking similarity between Boyle's and contemporary presentations of facts with sufficient stability, as he put it, to "make their own way," is the inscription of a discursive opposition between matters of fact and the speculations in which they are embedded. In Boyle's literary technology, there "were to be appropriate moral postures, and appropriate modes of speech, for epistemological items on either side of the important boundary that separated matters of fact from the locutions used to account for them: theories, hypotheses, speculations, and the like" (Shapin, 1996, pp. 66–67). For matters of fact "a confident mode was not only permissible but necessary" (p. 67). As for the experimental report's proper style for venturing speculations or hypotheses, or what Boyle calls "opinions," here is Boyle's advice to his nephew:

> In almost every one of the following essays I . . . speak so doubtingly, and use so often, *perhaps*, *it seems*, *it is not improbable*, and other such expressions, as argue a diffidence of the truth of the opinions I incline to, and that I should be so shy of laying down principles, and sometimes of so much as venturing at explications. (p. 67)

Boyle's distinction between a confident style for statements of fact and a hesitating style for speculative and interpretive statements is mirrored in the contemporary journal article's contrast between the interpretive problem setting of its introduction and the plain speaking of the methods and materials section and the reluctance to draw conclusions in the results and discussion (Knorr-Cetina, 1981). In both early modern and contemporary writing facts must be discursively stabilized. For Boyle the "separation of moral modes of speech and the ability of facts to make their own way were made visible on the printed page" (Shapin & Schaffer, 1985, p. 67). His " 'naked way of writing,' his professions and displays of humility, and his exhibition of theoretical innocence all complemented each other in the establishment and protection of matters of fact" (p. 69). Remarking on the rhetoric of the contemporary journal article, Knorr-Cetina notes that it "is well suited to the stereotyped image of science as presenting the 'facts' which others may use in making decisions" (1981, p. 123).

[15]On the reporting of circumstantial detail, Shapin and Schaffer write: "It is, however, vital to keep in mind that in his circumstantial accounts Boyle proffered only a *selection* of possible contingencies. There was not, nor can there be, any such thing as a report that notes *all* circumstances that might affect an experiment. Circumstantial, *or stylized*, accounts do not, therefore, exist as pure forms but as publicly acknowledged moves towards or away from the reporting of contingencies" (pp. 64–65; emphasis added).

In describing Boyle's objectifying project, Shapin and Schaffer ask: "If the obligation to assent to items of knowledge was not to come from human coercion, where did it come from?" The answer is the same today as it was in the seventeenth century:

> It was to be nature, not man, that enforced assent. One was to believe, and say one believed, in matters of fact because they reflected the structure of natural reality. . . . *Yet the transposition onto nature of experimental knowledge depended upon the routinization of these technologies and conventions.* (p. 79; emphasis added)

The continuity from Boyle's day to our own of the literary style of the "modest witness" is one of the most telling emblems of the necessity of such routinization. Facts must be inscribed in scientific writing so that they can "make their own way":

> The matter of fact can serve as the foundation of knowledge and secure assent insofar as it is not regarded as man-made. Each of Boyle's three technologies worked to achieve the appearance of matters of fact as *given* items. That is to say, each technology functioned as an *objectifying resource.* (p. 77)

Since the seventeenth century, "the objectivity of the experimental matter of fact [has been] an artifact of certain forms of discourse" (Shapin & Schaffer, 1985, pp. 77–78). Whatever other changes the experimental report has undergone in almost 350 years, many similarities attest to "modest witness" as an enduring literary style of scientific writing. Science's literary technology continues to construct the reader as a witness to a world of facts and phenomena of a natural world.

Contemporary studies of scientific writing show that the natural world represented in the journal article is not the same as the world of the laboratory. Knorr-Cetina shows how the scientific paper functions as an objectifying resource through its discursive construction of an alternate world. She calls this transformation of laboratory work through writing a "conversion of reason":

> We have observed a conversion into another currency, a transmutation into the totality of another language game. This conversion was itself a process. It started long before the paper was written, through the production of measurement data and other written traces of laboratory work, and continued with the collective enterprise through which these traces became caught, identified, and finally preserved within the double-threaded web of argumentation that distinguishes the finished paper. (Knorr-Cetina, 1981, p. 131)

The transformation is *from* the localized, contingent, opportunistic, highly situated, analogical, and practical reasoning governing laboratory resource selection, *to* the abstract, decontextualized space of the scientific paper. The transformation is at the same time a recontextualization, relocating possible decisions and possible conclusions on a stage of facts which "make their own way" in an objective, natural world, purged of all traces of human intervention.[16] The contingencies of actual scientific labor are transformed into an abstract, cognitive space, in which information from previous work and information produced in the laboratory are processed according to the rules of scientific rationality, thereby producing new information contributing to science's collective project of faithfully representing an objective, natural world.

Given the disequivalence between laboratory reason and its discursive reconstruction in the scientific paper, "the link between the laboratory and the scientific paper cannot be established by rules of cognitive transformation. The scientists who write a manuscript do not recall the research process and then proceed to summarize their recollections" (Knorr-Cetina, 1981, p. 130). Thus the paper is not a vehicle for the communication of information. Instead, the paper is a particular *discursive resource*, different from the laboratory's material setups, but no less an outcome of scientific labor. The erasure of particularity, situation, locality, and contingency represents the discursive fulfillment of its objectifying function. Since the paper stages a witnessing, not of the actual laboratory but instead of the "facts" of an objective, natural world, its witnessing is virtual in a double sense. Not only are the witnesses absent from the scene, but the scene itself is a discursive construct.

[16]Knorr-Cetina elaborates on this recontextualization: "In the transition from laboratory work to the scientific paper, the reality of the laboratory changed. We have seen the situationally contingent, opportunistic logic of research replaced by a generalized context of present and possible worlds, and the interest negotiations of particular agents transformed into a projected fusion of interests of technology, industry, the environment and a human population needing protein. We have seen the reasoned selectivity of laboratory work overruled by formulaic recitations of the doings which emerged from this selectivity, and the measured results of these doings purged of all traces of interdependency with their constructive creation. We have seen the indeterminacy of the laboratory reduced to the careful expression of scientific doubt which the paper allows" (1981, pp. 130–131).

The scientific paper's virtual witnessing stages a *simulacrum as witnessed*, through the process of which it becomes a phenomenon of the natural world. Insofar as knowledge of an experiment comes to depend upon its reconstruction by the scientific journal article, the particularity of the real laboratory situation is forever erased:

> The instrumental mode of production which results in laboratory measurements involves an almost total decontextualization, relieved only by the rationales found in the scientists' written notes. The literary mode of production which results in a published paper offers a recontextualization, but as we have seen, not one which brings back the memory of laboratory work. The transition is, at the same time, a conversion of the written traces themselves. Except in the memory of those who were present during the process, *it is an irreversible transition.* (p. 130; emphasis added)

The transformation of laboratory reasoning found in the scientific paper is typical of scientific resource conversion. Scientific work aims at launching resources used in other research contexts. The continuity of scientific practices does not arise from the logical coherence of an information space whose *telos* is the completeness of its representation of a natural world, but from the labor of resource conversion among scientific fields and resource extensions to transscientific fields. When the complex hybrids of the laboratory are taken up by others and used as resources in their own projects, "they undergo a recontextualisation and reconstruction similar to what we found in the writing of the paper" (Knorr-Cetina, 1981, p. 132).

Although the resource conversions of the journal article are characteristic of scientific products generally, there are also important differences. A literary technology's products are *discursive* objectifying resources. To standardize and routinize discursive decontextualizations and recontextualizations through a discipline of scientific writing creates the objectifying resources for the discursive construction of objectivity. In other words science's literary technology creates resources for the articulation of objectivity, nature, scientific truth, and scientific knowledge. Formal writing is crucial to establishing the documentary techniques for the institutionally authorized enunciation of scientific truth. Studies of scientific practices therefore imply that the journal article is central to such practices, not because it conveys information but because of the centrality of objectifying resources to the cultural phenomenon we know as natural science.

Acknowledgments

The author gratefully acknowledges the support of grant R0963A01 from the Social Sciences and Humanities Research Council of Canada for the research presented here.

References

Anderson, F. H. (1948). *The philosophy of Francis Bacon*. Chicago: University of Chicago Press.

Bacon, F. (1960). *The new Organon and related writings*. F. H. Anderson (Ed.). New York: Liberal Arts Press.

Bazerman, C. (1983). Scientific writing as a social act: A review of the literature of the sociology of science. In P. V. Anderson, R. J. Brockmann, & C. R. Miller (Eds.). *New essays in technical and scientific communication: Research, theory, practice*. Baywood's Technical Communications Series, vol. 2. Farmingale, NY: Baywood Publishing.

Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison, WI: University of Wisconsin Press.

Bazerman, C. (1994). *Constructing experience*. Carbondale, IL: University of Illinois Press.

Bernal, J. D. (1939). *The social function of science*. London: Routledge.

Callon, M., Law, J., & Rip, A. (1986). *Mapping the dynamics of science and technology: Sociology of science in the real world*. Houndmills, Basingstoke, Hampshire; London: MacMillan Press.

Collins, H. M. (1992). *Changing order: Replication and induction in scientific practice*. Chicago: University of Chicago Press.

Fuller, S. (1995a). Cyberplatonism: An inadequate constitution for the republic of science. *The Information Society 11*, 293–303.

Fuller, S. (1995b). Cybermaterialism, or why there is no free lunch in cyberspace. *The Information Society 11*, 325–332.

Garvey, W. D. (1979). *Communication: The essence of science. Facilitating information exchange among librarians, scientists, engineers and students*. Oxford/New York: Pergamon Press.

Hacking, I. (1983). *Representing and intervening: Introductory topics in the philosophy of natural science*. Cambridge: Cambridge University Press.

Hagstrom, W. O. (1965). *The scientific community*. New York: Basic Books.

Haraway, D. J. (1996). Modest witness: Feminist diffractions in science studies. In P. Galison & D. J. Stump (Eds.). *The disunity of science: Boundaries, contexts, power*. Gen. eds. T. Lenoir & H. U. Gumbrecht. Stanford, CA: Stanford University Press.

Harnad, S. (1995a). The PostGutenberg galaxy: How to get there from here. *The Information Society 11*, 285–291.

Harnad, S. (1995b). Sorting the esoterica from the exoterica: There's plenty of room in cyberspace. A response to Fuller. *The Information Society 11*, 305–324.

Hayes, R. M. (1992). The needs of science and technology. *Science and Technology Libraries 12*(4), 3–33.

Knorr-Cetina, K. D. (1981). *The manufacture of knowledge: An essay on the constructivist and contextual nature of science*. Oxford: Pergamon Press.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Cambridge, MA: Harvard University Press.

Latour, B., & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. 2d ed. Intro. by J. Salk. London and Los Angeles: Sage Publications.

Martin, J. (1992). *Francis Bacon, the state, and the reform of natural philosophy*. Cambridge, MA: Cambridge University Press.

Merton, R. K. (1973a). Behavior patterns of scientists. In N. W. Storer (Ed.). *The sociology of science: Theoretical and empirical investigations* (pp. 323–342). Chicago: University of Chicago Press.

Merton, R. K. (1973b). The normative structure of science. In N. W. Storer (Ed.). *The sociology of science: Theoretical and empirical investigations* (pp. 267–278). Chicago: University of Chicago Press.

Merton, R. K. (1973c). Priorities in scientific discovery. In N. W. Storer (Ed.). *The sociology of science: Theoretical and empirical investigations* (pp. 287–324). Chicago: University of Chicago Press.

Merton, R. K., & Zuckerman, H. (1973). Institutionalized patterns of evaluation in science. In N. W. Storer (Ed.). *The sociology of science: Theoretical and empirical investigations* (pp. 461–496). Chicago: University of Chicago Press.

Mikhailov, A. I., Chernyi, A. I., & Giliarevski, R. S. (1984). *Scientific communications and informatics*. R. H. Burger (Trans.). Academy of Sciences, State Committee on Science and Technology (VINITI). Arlington, VA: Information Resources Press.

Miksa, F. L. (1998). *The DDC, the universe of knowledge, and the post-modern library*. Albany, NY: Forest Press.

Pickering, A. (Ed.). (1992). *Science as practice and culture*. Chicago: University of Chicago Press.

Pickering, A. (1995). *The mangle of practice: Time, agency, and science*. Chicago: University of Chicago Press.

Price, D. J. de S. (1970). Citation measures of hard science, soft science, technology, and nonscience. In C. E. Nelson & D. Pollack (Eds.). *Communication among scientists and engineers* (pp. 3–22). Lexington, MA: Lexington Books, D. C. Heath & Co.

Price, D. J. de S. (1986). *Little science, big science . . . and beyond*. New York: Columbia University Press.

Royal Society (Great Britain). (1948). *The Royal Society Scientific Information Conference, 21 June–2 July 1948: Report and papers submitted*. London: Royal Society.

Shapin, S. (1996). *The scientific revolution*. Chicago: University of Chicago Press.

Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life*. Princeton: Princeton University Press.

Toren, N. (1983). The scientific ethos debate: A meta-theoretical view. *Social Science and Medicine 17*(21), 1665–1672.

The Game of the Name: Nomenclatural Instability in the History of Botanical Informatics

Geoffrey C. Bowker

Abstract

Botanical nomenclature of vascular plants dates back to the first edition of *Linnaeus* in 1753. Since then a series of attempts have been made to deal with the problem of name stability—how to ensure that a given species will have the same name all over the world and over time. In the 1820s the Kew rule was developed out of Kew Gardens to deal with priority in generic names; in the 1860s George Bentham worked on two large projects to stabilize systematics. In the 1890s the Berlin rule (limited priority for names that had either fallen into disuse or never been accepted) came into conflict with the Philadelphia rule (according to which priority was absolute). Throughout the past century a series of international conferences have dealt with issues arising from these controversies.

The principal issues are twofold. On the one hand, it is highly desirable to be able to change the names of plants when new scientific insights come into place. On the other hand, it is extremely difficult and costly to change the names of plants. In this paper I explore the developing positions with respect to name stability, first, with respect to the nature of scientific cooperation (here in the form of the establishment of international conferences to deal with questions surrounding the issues and the development of agreements about which scientific journals could carry new names), and second, with respect to the development of information technology (computer technology, for example, makes name flexibility in some ways easier to propagate). I argue that the issue of name stability has been the site of a series of significant discussions about the nature and storage of scientific information. This premise forms the basis for a recent work of mine about the history of classification systems in medicine (Bowker & Star, forthcoming), in which I have argued that work done at the deep infrastructural level of classification and nomenclature systems is closely tied to both information technology developments and organizational histories (the organization of the profession of medicine or botany in these cases). I also argue in this paper that it is impossible not to encode deeply into the information infrastructure some specific readings of the state of knowledge and of the state of relations between often competing professional groups (systematists, botanists, farmers) and that this has had significant consequences both for knowledge production and organizational change.

Introduction

A banner headline in the *Independent* for 23 November 1998 reads "Scientists Reclassify All Plants." The headline is wildly inaccurate: Reclassification is a long, slow process, and there is no simple path from the molecular sequencing techniques referred to in the body of the article to the development of new plant classifications. Further, even in the world of electronic databases we are moving into, no touch of a button would allow us to usher in a new system. On the contrary, when a given database of plants, of the ecology of a given area, of paleontology, and so forth is designed, it necessarily draws on a contemporary classification and will rarely be updated (and will be difficult to update) should the classification change. The result is a tower of Babel, where numerous outdated classifications present themselves to the scientific researcher with equal force: Indeed they must be used if the associated data is to have any value. In this paper I look at the two-century-old effort to establish and stabilize scientific names for plants and argue that the attempts have been so difficult to make work because precise boundaries for priority (who first named the plant), publication (where the name is published), and reach (who has the authority to name) are integral questions about the organization of work in systematics and about the scientific features of a given plant. I will draw attention to issues that have arisen with the need for long-term, wide-scale information storage and retrieval (cf Bowker & Star, 1994, forthcoming) and will discuss the range of solutions that have been worked out over time—and how these affect the ordering of knowledge in the field of botany.

Diana Crane (1972, p. 8) claims that in scientific literature "the 'life' of a paper is very short, with the exception of a few classics. Papers published five years ago are 'old.' Papers published more than fifteen years ago are almost useless in many scientific fields." In this paper I will examine a field of science in which this is emphatically not the case—the field of botanical nomenclature. I will discuss the issues that have arisen over the past 250 years as botanists have tried to develop universal, standard names for plants. The practice of botanical nomenclature is not fully aligned with the practice of botanical classification: Although in principle Linnaeus's system is both classificatory and nomenclatural, in practice many names are retained beyond their classificatory currency.

Crane's model works best in physics, where no assumption is made that information collected in the early nineteenth century will still be of interest to the current generation of field theorists. There is the assumption (Poincaré, 1905, for example) that new theories will reorder knowledge in the domain effectively and efficiently. And since Kuhn (1957), most would accept that a major paradigm change in the understanding of gravity, for example, renders previous work on incline planes literally incommensurable, not to mention technical improvements making the older work too imprecise. Chemists have somewhat more need to delve into older material (but certainly not in the issue of naming chemicals—the procedure has been internationally standardized since the mid-nineteenth century). Astronomers trawl back further in time, seeking traces of supernovae in ancient manuscripts. But they are just as likely to look at monastery records as at Tycho Brahe's original data.

In order to name plants, botanical taxonomists consistently need regular reference to scientific literature dating back to the mid-eighteenth century. Botanical nomenclature of vascular plants dates back to the first edition of Linnaeus's code in 1753. Since then a series of attempts have been made to deal with the problem of name stability: how to ensure that a given species will have the same name all over the world and over time. In the 1820s the Kew rule was developed out of Kew Gardens to deal with priority in generic names; in the 1860s George Bentham worked on two large projects to stabilize systematics. In the 1890s the Berlin rule (limiting priority for names that had either fallen into disuse or never been accepted) came into conflict with the Philadelphia rule (according to which priority was absolute). Throughout the past century a series of international conferences have dealt with these issues.

The principal problems are twofold. On the one hand, it is highly desirable to be able to change the names of plants when new scientific insights are developed. On the other hand, it is extremely difficult and costly to change the names of plants. Consider the common tomato. Recent systematics research has suggested that its genus *Lycopersicon* should lose its status and that the plant should be accreted once more into the genus *Solanum*. At the time a recent book on nomenclatural stability (Hawksworth, 1991) was produced, the International Seed Trade Association was protecting the old name until the systematics debates were completed. But within the Botanical Code (affecting scientific publications, for example) the name could change during such discussion (Brandenburg, 1991). Changes in name introduced as taxonomic theory develops can have large-scale economic consequences: It could cost tens of millions of dollars to relabel packets of tomato seeds, revisit regulations, and so forth. One commentator noted that "single name changes can cost the horticultural trade millions of dollars, and . . . nurserymen would go out of business if they took the matter seriously" (p. 30). Brandenburg asks:

> Have you ever tried to explain to a nurseryman, a plant trader, or a customs officer which name he should use for the tomato? Can you imagine the reaction if such a name has to be changed three times in ten years? If you have done so, you have a perfect explanation for the unpopularity of plant taxonomy amongst those in any work related to agriculture, horticulture, and silviculture: the inability of plant taxonomy itself to settle discussions with only a nomenclatural background. (P. 24)

Anderson's (1991, p. 96) account raises the problem that even if a decision is made, the confusion will remain:

> Accumulating evidence suggests that *Lycopersicon* may not stand as a genus distinct from *Solanum*. If they are combined there is no chance that *Solanum* will be dumped for *Lycopersicon*, so even though the name *L. esculentum* has been conserved, if the tomato is treated as a *Solanum* its name will become *S. lycopersicum*, which will provoke howls of anguish from agriculturalists and other non-taxonomists. We could conserve the name *S. esculentum*, which would go halfway toward solving the problem, but that would still leave our critics rabid. And there will be no escape, because that name-change will be essential to reflect accurately that opinion about the

> relationships of the tomato. Worse, because reasonable people may differ in this matter, some taxonomists will continue to use *L. esculentum*, some will use *S. lycopersicum*, and the matter could continue unresolved for many years.

The issue of different interests between the person in the field or in the marketplace and the systematist who ultimately arrogates the right to name appears again in the case of the Douglas fir:

> Proposals to establish a list of *nomina specifica conservanda* have been presented at several successive congresses. Such lists tend to be favoured especially by foresters and agronomists, who are not much concerned with formal taxonomy. I can understand the outrage of foresters at the change of the name of Douglas fir from *Pseudotsuga taxifolia* to *P. menziesii*, all because of some bibliographic digging by a man who knew little about the tree. (Cronquist, 1991, p. 302)

In the case of tomatoes we have differences between the scientist who knew too much and the practical field person; here we have anger directed at the armchair scientist who knows too little. The language is extremely strong in both cases, with words like *outrage* and *anguish* being used as well as reference to "rabid" critics. The authority to name clearly is an issue that touches some raw nerves among those who feel that names are being pushed upon them. All this century—since the Vienna Botanical Congress of 1905—some names have been preserved against the ravages of the taxonomist.

What is happening here in informational terms applies over a range of disciplines and can be expressed simply—even if over one hundred years of international meetings have not come close to settling the problem. In order to keep track of results in the sciences, you need to be sure of what you are dealing with—a rose should be a rose should be a rose, whether in seventeenth-century Leipzig or twentieth-century Pesotum. So the first principle is creation of as much name stability as possible. New understandings of plants, however, can lead to rearrangements of taxonomy: Two genera seen as historically distinct—the Chinese cabbage and the European turnip, for example—might now be seen as one (Chauvet, 1991). Both vegetables are of considerable economic importance, so it would be very difficult to change the name. Further, it would require painstaking indexing work to track the losing genus through the literature (it is akin to the problem in history of following women through name changes on marriage). This merging and splitting work is done at every level—from the species up to the kingdom (at which level the disputes between the Zoological and Botanical Codes about who gets to name, say, candidates for the kingdom of Protista [in which the members can be considered plants or animals] are called ambiregnal problems). But if the names are not changed, then the naming system loses its connection with theory and becomes more of an arbitrary mapping of the world, which militates against the whole point of producing names in the first place. To make matters worse, this area is far more intense in its demands than some others. In medicine, where exactly the same theoretical issues arise with the International Classification of Diseases—actually a statistical classification and a nomenclature (Bowker & Star, forthcoming)—we are dealing with far fewer entities than the number of plant and animal species.

The Time of the Name: Priority

Over time the botanical community has developed a set of strictly defined bureaucratic procedures for dealing with naming. One of the basic principles is priority—often a bugbear in scientific communities but here a particularly difficult issue. The problem is that in a widely distributed literature—across all continents and a number of disciplines each with their own sets of journals—a rule is needed to decide which name should be standardized on. Priority has its own meta-priority as the solution of choice to such problems: This can lead, however, to standardizing to a name that is in minority use and thus losing an almost universal name.

The principle of priority is rooted in the work of Linnaeus, who invented the binomial naming system, which provides a consistent mode of naming all plants. In general, it has operated on the principle that a plant name be recognized if it "has been given after 1753; is in accord with the Linnaean system (a thorny issue since many non-Linnaeans in the early nineteenth century gave names that nevertheless can be interpreted as of binomial genus-species form); and is not invalid according to the current rules of botanical nomenclature" (IRBN, 1935).

A plant's full name evokes a detailed history, after the binomial giving an abbreviated reference to the author of the work where the name first occurred and a reference to the site of the publication.

This apparently simple set of criteria is notoriously hard to carry out in practice. One needs rules that are rigid enough to allow an unambiguous determination of difficult cases, yet flexible enough to accept publications that do not strictly follow the rules but even so have led to a universally accepted name. The first rule that was developed to standardize naming was the so-called Kew rule, applied by botanists at Kew Gardens in London in the 1820s. According to this rule:

> Only epithets that are already associated with a generic name are considered from the point of view of priority when that genus is being revised—this is the so-called Kew Rule. Priority dates from the time that the specific epithet is first associated with the generic name. Hence epithets that may be older, but which have only been associated with species placed in other genera, can be ignored, and when genera are combined, well-established names are less likely to be changed. The major issue in the Kew Rule is how priority is interpreted when genera are combined, a minor, but associated issue is that of whose names are to be cited when a plant name is transferred. (Stevens, 1991, p. 157)

The epithet here is the distinguishing name for a species, often describing one of its features. Thus the epithet "esculentum" in *L. esculentum* (the old name for the tomato) means "edible." The major effect of the Kew rule was to give botanists some flexibility in renaming; if a plant was given a name under what was now believed to be the wrong genus, then that name did not have priority. Priority only accrued if genus and species remained constant.

In the 1860s there were two large-scale projects to standardize naming: George Bentham's *Genera plantarum* and Alphonse de Candolle's Laws of Botanical Nomenclature (Stevens, 1991). Bentham strongly defended the Kew rule and argued that there should be room in the canon for names that were not strictly correct in Linnaean terms but that were nevertheless widely accepted (Stevens, 1991, p. 162).

De Candolle's work included lengthy discussions on the nature of priority, which drew particular attention to the tension between the name as history and the name as signifier. His laws formed the kernel of the current rules of nomenclature, which were maintained through a series of international biological congresses held throughout this century. He drew attention to issues of just how much history could be wrapped up in a name. Thus he cited Friedrich Kirschleger's discussion of *Mulgedium alpinum L. sp. 117 (sub: sonchus), Less. Syn 142*. The "L." here refers to Linnaeus, who was not the namer of the genus *Mulgedium*—in fact the epithet *alpinum* was Linnaean and the full name came from Christian Lessing—named after the first parenthesis. De Candolle pointed out that this long and complex name would frequently be shortened in card catalogs or other lists to *Mulgedium alpinum L.*, which would give a false history but would still be an effective, unambiguous name. He argued in two ways that the history wrapped into the name did not serve the purpose of providing glory, but was rather a simple convenience for arriving at an identifier:

> When one wants to pay homage to a botanist, one dedicates a genus to him. When you want to speak about his merits or demerits on the subject of a given species or genus, one adumbrates and discusses his opinions either in the text of a description or by some parenthesis in a synonymy—however the citing of a name or names in a plant name does not in itself express either merit or demerit. It is the statement of a fact—that is to say that such and such an author was the first to give such and such a name to a genus, or that he was the first to attach this species to that genre. (De Candolle, 1867, p. 47)

He argued that "the name is what counts most. . . . You might change whatever you want in your description of the genus *Xerotes*, Br.; but one thing is fixed and certain and that is that Brown, in 1810, designated a genus with this name" (p. 53). He pointed out that the traveler who first picked the plant should perhaps be rewarded—but that priority in publication was the naming rule (p. 54). De Candolle even looked forward to a future day when the current set of names might drop out—once science had succeeded in describing definitively what plants there were on the face of the earth, then the current "scaffolding," which contained many local exceptions to strict rules, might fall away. This, however, was not for the immediate future (p. 7). Here he is echoing Auguste Comte on scientific classifications, and indeed the French revolutionary calendar on the dating of events: When scientific precision is introduced, the history can drop away.

The flexibility advocated by Bentham and de Candolle enraged Otto Kuntze. In 1891 he produced a list of thirty thousand names that would have to be changed under a strict application of the laws of priority, and he wanted these changes to be effected (Briquet, 1906, p. 5). He excoriated proponents of the Kew rule, and called Bentham in particular "a great sinner in nomenclature" (Stevens, 1991, p. 162). Though many admired his work in principle, its root and branch changes were considered impractical in general. They led directly to the Berlin rule, according to which generic names not in general use fifty years after their publication could be abandoned. In 1906 Kuntze stormed out of the International Botanical Congress in disgust, claiming that his protests were not being taken seriously enough (Briquet, 1906, p. 112).

Jean Briquet was a prime mover at the International Botanical Congress in Paris in 1900, and he headed the commission set up by that congress to determine a new code of botanical nomenclature. Priority remained central, but it was palliated by the conservation of certain names that lacked priority but had universal acceptance. Needham commented in 1910 on the work of the commission:

> We have accepted the alteration of hundreds of well-known names that are root-names of many more genera within their respective groups: and such derived names, once of great assistance to the memory, have, so to speak, the props knocked from under them.
>
> Finally, and most lamentably of all, by our hasty and profitless abandonment of even the best-known family names we have broken with our best traditions . . .
>
> The pursuit of stability through rules of priority that has led to all this is surely one of the most singular of contemporary psychological phenomena. (1910, p. 296)

We shall return to Needham's reference to memory in the section on euphony below. What is significant here is that priority is seen by Needham as an unnecessary principle indicative of, if anything, psychological disorder:

> Why should it [the international commission] determine merely whether a certain forgotten name, abandoned by its author and never used, is really eligible for use under the rules of the code? It grieves me to see fifteen big brainy men, capable of doing something rational, put into a hole where they are expected to do only such little sinful things as this. (1910, p. 296)

The word *sinful* is evocative of the passion of the priority debate, echoing as it does Kuntze's charge against Bentham.

Throughout this century the International Congresses have ever further refined in parallel the application of the rules and the granting of exceptions to them. A typical entry in a relatively recent (1965) proposal for the conservation of a name gives some idea of the kind of work—at once botanical and bibliographical—that is involved in the maintenance or breach of priority. B. Verdcourt from Kew proposed conserving the generic name *Warburgia* Engl., 1895 against *Chibaca* Bertol. F., 1853 (Cannellaceae). He argued that *Chibica* was considered invalid soon after its publication and that Bentham and Hooker had added it to their "genera ramanent indefinita et nomina delanda" (genera that remain undefined and deleted names). In an Italian journal in 1937 Emilio Chiovenda suggested that *Chibica* was a member of the family Canellaceae and was identical with *Warburgia breyeri* Pott. Chiovenda had not actually seen any *Warburgia*-type specimens, however, so Verdcourt held fire. Then in 1964, while he was passing through Bologna, Verdcourt searched for Chiovenda's type specimen. But it could not be found. German troops lodging at a farmhouse near the herbarium had burned the collection. Verdcourt went back to the literature with his interest piqued and determined that *Chibica* was *Warburgia* but that the latter should be conserved since the former name "is virtually unknown and has never been used in any flora or paper other than that by Chiovenda" (Nomina Conservanda, 1965, pp. 27–28). Thus, in order to effectively breach the principle of priority, Verdcourt had to prove that priority should apply according to the rules of nomenclature and then make the argument that in this case it would cause unnecessary problems to actually apply it. Equally, when the rule is applied, it is often necessary to do intense bibliographical work. Galtier (1986, p. 6) notes that it was necessary now to get down to the day and month of publication in many cases.

Priority, then, has been seen by some as everything from a pure naming convention (De Candolle's position) to a matter of grave importance (Kuntze). The list of botanical names willy-nilly serves as both an honor roll (and in so doing necessarily contains a highly formalized and abbreviated account of the history of each name) and as a set of arbitrary identifiers (since everything must be called something). There was a short-lived attempt to introduce the concept of "numericlature" (giving each taxon a universal number rather than a name), but it did not gain many adherents (Little, 1964). Names are taken much more seriously by scientists and by the general public than numbers or arbitrary identifiers, so that rather than solving the problem of naming it would merely have added another layer to its complexities (cf Bowker & Star, forthcoming, Chapter 2 on alternative naming schemes for viruses).

The Space of the Name: Effective Publication

The issue with priority is who came first; the issue with publication is from which publication did they come? In the early nineteenth century, when there were fewer scientific journals, the most general problem was dealing with works not in English, French, or German. Over time the number of journals has increased dramatically, and so the amount of bibliographical work that must be done to locate and propagate a name has risen in conjunction (Kirk & Cannon, 1991, pp. 279–280).

De Candolle's first principle of botanical nomenclature, accepted by Briquet into the International Code and still in place is that "Natural History can make no progress without a regular system of nomenclature, which is recognized and used by the great majority of naturalists in all countries" (De Candolle, 1867, p. 13; Briquet, 1906). He discussed the problem of referencing publications within plant names. With that passion for system that characterizes much work in this field, he discussed the problem of author abbreviations in botanical names, and enumerated some forty-seven vowel and diphthong combinations that could hide in between the "h" and the "k" in "Hkr:" and then pointed out that the same forty-seven could hide between the "k" and the "r," leading to 2,209 possible names (De Candolle, 1867, p. 56). Just as he argued that the name of the plant was just a name and not an attribution of glory, so he argued that the publication of a name was just a publication and was not something still owned by its author:

> Can an author who regrets having published a name change it? Yes, but only in the those cases where the names could by changed by any botanist. In effect the publication of a name is a *fact* that the author cannot revoke. (De Candolle, 1867, p. 57)

De Candolle did not, however, discuss just what a publication was.

By the time of the Paris Congress in 1905, the definition of *publication* had become an important issue—complicated of course by the fact that the further one went back in time the less well defined was the field of scientific publication. (I note in passing that many of the rules adopted by botanists with respect to nomenclature can be read as an attempt to apply retrospectively whatever the current canons of scientific publication were to previous generations. This inevitably led to distortions of the historical material and so to a kind of active reading in science that would only be developed in literary criticism in the mid-twentieth century—a movement countered, for example, by De Candolle's enunciation of the principle of "never making an author say what he has not said" [Stevens, 1991, p. 159].) Briquet's commission proposed the definition that "Publication is effected by the sale or public distribution of printed matter or indelible autographs. Communication of new names at a public meeting, or the placing of names in collection or gardens open to the public, do not constitute publication" (Briquet, 1906, p. 53).

Equally thorny at that period was the question of whether "diagnoses" (formal descriptions) of plants had to remain in Latin. This problem was raised by the Spanish delegates, who wanted their language accepted alongside French, German, English, and Italian as legitimate languages for a diagnosis (International Botanical Congress, 1906, p. 131). They lost the battle, on arguments such as that by Professor Maire that

> The principle of an obligatory Latin diagnosis . . . is the only means of conserving at present an international language, which language is an immense privilege for systematic botany. If we admit diagnoses in three modern languages, then everyone else will want to join in: after the Chinese there will be no reason to refuse the Papuans, the American Indians and all peoples who may one day accede to scientific life. Systematic botany would become a veritable Tower of Babel. (Fifth International Botanical Congress, 1931, p. 583)

There is an element of irony in the enforcement of Latin in the interests of internationalism, but it is unclear what alternatives existed. Many botanists were concerned about the discovery of plant descriptions in valid form in Russian, for example, supplanting (to coin a phrase) current names in Western Europe. This could be particularly difficult if different philosophies of naming were in operation. It has been suggested, for example, that in Marxist Russia there were no infraspecific categories because these were not acceptable to dialectical materialism's insistence on the irreducibility of species (Heywood, 1991, p. 54; cf Graham, 1972).

Over the course of this century the issue of how many journals to look in has been problematic. At Kew Gardens currently some seven hundred journals "are regularly scanned, as well as monographs, floras and other works in which new names might be found" (Lock, 1991, p. 287). Each new discipline that has grown up has spawned new journals. Fossil species have long presented difficulties. For example, the *Journal of the Geological Society of London* is only rarely read by neobotanists (those concerned with current flora), and yet paleobotanists have proposed new fossil genera in it (Boulter, Chaloner, & Holmes, 1991, p. 238). From 1 January 2000 "the names of newly described botanical (including fungal) species will have to be registered in order to be validly published," and the Clearing House Mechanism is being adapted to coordinate this on an international basis (Heywood, 1991, p. 57).

The issue with publication therefore has been how to be sufficiently universal so as to accept all scientific work done throughout the world and yet sufficiently restrictive so as to make the information management problem tractable. The latter criterion has been frequently met by regularly underrepresenting work not in English or some other major European language and ignoring work not in an ill-defined set of central journals.

The Sound of the Name: Euphony

Euphony has been a surprisingly resilient problem in the history of the naming of plants. Linnaeus has as one of his basic recommendations that plants, in order to be easy to remember—all botanists must know and remember all the genera (Cain, 1958, p. 144)—should be easy to pronounce. That is, they should be euphonius. George Bentham observed in 1838 that "names that seemed very difficult for an Englishman to pronounce might be easy for a Pole, Russian or German, and vice versa" (Stevens, 1991, p. 160). By the next century, however, euphony was back on the agenda. In the 1905 Vienna Congress (one of the turning points in the history of botanical nomenclature), Linnaeus's recommendation might be seen to be echoed in the following:

> V. Botanists who are publishing generic names show judgement and taste by attending to the following recommendations: . . .
> c) Not to dedicate genera to persons who are in all respects strangers to botany, or at least to natural science, nor to persons quite unknown.
> d) Not to take names from barbarous tongues, unless those names are frequently quoted in books of travel, and have an agreeable form that is readily adapted to the Latin tongue and to the tongues of civilized countries. (Briquet, 1906, p. 39)

Each line requires a little elaboration. The "judgement and taste" phrasing is there to emphasize that this is a recommendation and not a requirement; the congress was attempting to deal with the problem of consistently naming all taxa worldwide for all time, and so wanted to keep requirements to a minimum. I have included point "c," which is not a principle of euphony but does give an indication of the company that euphony kept: adjurations to civilized behavior in contemporary terms and ways of excluding the outsider and the underdeveloped in more recent coinage. As late as 1971 a new botanical nomenclature (NBN) was proposed that would preserve euphony in similarly ethnocentric fashion. The NBN uses Esperanto, where "the words are pleasing to the ear, there is enormous flexibility in word-formation, etc." (De Smet, 1991, p. 180).

Thus one person's euphony is another's cacophony, and yet, as with priority and publication, it is a *prima facie* reasonable requirement. The problems have arisen—again as with the others—when you try to turn a set of precepts that have worked for a loosely defined club of largely Western European natural philosophers into a system that can work universally.

The Reach of the Name: Organizational Dimensions

The naming system in botany has served as a means of demarcating professional and research communities one from the other. De Candolle in 1867 made clear the distinction between botany and zoology:

> [Linnaeus's system] has often been cited in philosophy courses. It has been considered superior to chemical nomenclature, because it lends itself better to changes necessitated by progress. Botanists professed a veritable cult for the system. They prided themselves on having better understood and developed it than the zoologists. (P. 3)

Indeed his first principle of botanical nomenclature, taken up in the first international code and still the first principle to this day, was that "Botanical nomenclature is independent of zoological nomenclature, in the sense that the name of a plant must not be rejected merely because it is identical with the name of an animal." (ICBN, 1956, p. 12).

This distinction has led to a series of border disputes concerning just what should or should not be included in the nomenclature. One thorny issue has been ambiregnal species—species that might equally well be designated plant or animal. This problem has been on the rise with new phylogenetic work "increasing the number of major new inherently ambiregnal clusters of autotrophic (plant-like) and heterotrophic (animal-like) protists" (Patterson & Larsen, 1991, pp. 197–198). Most believe that generating a single code is just not going to happen, but that arbitrarily assigning protists to one code or the other is equally problematic (p. 201).

A dispute erupted with bacteriologists in the Fifth Botanical Congress. The bacteriologists, led by R. E. Buchanan, wanted an exemption from the need to use Latin in their diagnoses of specimens. Botanist Thomas Haumann came back with the argument that what "bacteriologists, doctors and chemists call a 'description' " is not what botanists would call one. He argues that "the bacteriologists have not yet reached the stage of development which would permit the establishment of an accord between their still rudimentary systematics and a rational systematics." Buchanan responded that bacteriologists were doing serious and rational work and that "if the congress cannot accept my motion, the bacteriologists will separate themselves from the botanists and will develop their own rules" (Fifth International Botanical Congress, 1931, pp. 588–590). And the bacteriologists subsequently did break off from the botanists—underscoring the move with a decision in the 1960s to abandon priority and free "themselves of the burden of past names and literature by adopting a list of all bacterial names in use, removing from nomenclature all names not listed, and adopting a process of registering all new names proposed henceforth" (Ride, 1991, p. 106). This is not an isolated instance in scientific communities. Indeed, a similar dispute occurred this century between plant and animal virologists and was exacerbated by the discovery that some viruses could jump between plants and animals. The disagreement led to an enforced merger of two fiercely different codes, with the proponents of zoological and botanical nomenclature thundering dismissals of the others' system (Matthews, 1983; Bowker & Star, forthcoming).

Currently there is a whole apparatus of name protection that echoes the apparatus of species protection. Thus, when a plant goes from being a weed to a useful variety, its name goes from being changeable to being fixed:

> In cases where names of economically important species are involved, the Code provides for conservation in order to preserve current usage. Judging from the success of past species conservation or rejection proposals, it is not always clear as to what constitutes an economically important species. Some weeds of significant agricultural importance have not qualified for name conservation. With the expanding potential of gene transfer for crop improvement involving more distantly related taxa, more species will become useful to agriculture. Communication about such species depends on a stable nomenclature. (Gunn, Wiersema, & Kirkbride, 1991, p. 18)

In general, two options are open for a given name: It can be "conserved," which means that it is protected indefinitely against the ravages of taxonomy and nomenclatural reasoning, or it can be "stabilized," which means that it will be protected for a given period of time while debate proceeds. As indicated in the introduction, serious economic consequences can flow from decisions made by taxonomists. The problem of cultivars (cultivated varieties) and their naming has been a constant one: It has been remarked that botanical "snobbery" has meant the overlooking of "substantial horticultural works of the late nineteenth and early twentieth century" (Hawksworth, 1991, p. 106) as published sources for priority purposes.

Not only the farmers and nurserymen have problems with the current politics of naming. Consider the fossil, that apparently most useful of traces of the past, which is currently being massively underused partly because of naming problems:

> The use of fossils has in recent years come to be regarded as cumbersome and unproductive; the work is said to abound with unimaginative complacency, with the obscurity of esoteric terminology, and with lack of compatibility of treatment of different groups including even that between the fossils of plants and animals. As a result, much effort has been directed by geologists in solving their stratal problems towards employing any other available physical or chemical phenomena, and thus to avoiding altogether the "expensive" and supposedly ineffective use of paleontologists and their fossils.
>
> Paleontologists, who almost all continue to believe that their fossils and the distribution of these form the only viable method of discriminating diverse and confusing strata, are striving to present their fossils more ingeniously and to win back the confidence of the geologists. (Hughes, 1991, p. 39)

This has led to a situation where geologists from some petroleum companies have abandoned the scientific literature and developed their own coding for fossil remains. Even when the agreed-upon naming procedure is followed, the fossil genera might be buried in, say, the *Journal of the Geological Society of London* (p. 238), where no database manager or neobotanist would look for a new genus. There is a set of difficulties associated with fossil use that compounds these problems. As one author notes, there are just too many conflicting uses for them: "Many of the difficulties in palaeobotany and its sub-discipline, palynology, occur as a result of the sometimes conflicting aims of botanical and geological researchers handling these data" (Boulter, Chaloner, & Holmes, 1991, p. 232). Further, fossil plants are rarely complete so that different parts of a plant will be classified differently; and the mode of fossilization—permineralization or compression are two main modes—often gives rise to different classifications, unless a linkage can be made through a contingent Rosetta stone (Galtier, 1986).

The situation today is in some ways much the same as it has been over the past two hundred years. A recent author (Klemm, 1990) noted that

> For any given conserved tropical wildland we are confronted with a problem roughly analogous to receiving an enormous library with no call numbers, no card catalogue, and no librarians—and the library being in a society that is only minimally literate and not even certain that reading has much to offer. The library is hardly more than highly flammable kindling in such a scenario. (P. 23)

Indeed renaming in this context can be highly problematic. When subspecies or varieties are elevated to full species rank,

> This may have very unfortunate consequences from a legal point of view when the species to which the subspecies or variety belonged before the nomenclatural change is listed as a protected species, the result of the split is that the new species loses its protected status unless the legislation is amended to add it to the list. (P. 33)

The move to register all names, to agree on model data structures and formats for biological databases in order to facilitate biodiversity management (Heywood, 1997, p. 12) is just as urgent and just as overly optimistic as the calls of De Candolle for a rational system of nomenclature.

Conclusion: What's in a Name?

In this paper I have endeavored to show that the production of consistent names for all plants is a very rich organizational and intellectual process. Over time it has involved setting up rules for the reading of documents—and indeed producing an understanding of just what kind of activity reading is; deciding just what kind of a thing a publication is; endeavoring to find a name with a pleasing, memorable sound (and to deal with cross-cultural issues in deciding euphony); and negotiating with other scientific groups (zoologists, bacteriologists) and with commercial and regulatory bodies (horticulturists, nurserymen, government agencies).

This set of issues is matched by other bodies (for example, epidemiologists) who try to maintain datasets for extended historical periods and geographical sweep. They are issues that each generation has generally ignored as the new set of information technologies is brought into play along with its particular dream of a common language (Rich, 1978).

Naming is a difficult thing to do. It is a site of important decisions for the organization of knowledge and for the organization of scientific work, and it is an activity with political and economic consequences. These dimensions should be fully factored into the development of new information systems to deal with the burgeoning huge datasets that are a necessary adjunct to biodiversity management.

References

Anderson, W. R. (1991). Should we change the code? Concerns of a working taxonomist? In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 95–103). Konigstein, Germany: Koeltz Scientific Books.

Boulter, M. C., Chaloner, W. G., & Holmes, P. L. (1991). The IOP plant fossil record: Are fossil plants a special case? In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 231–242). Konigstein, Germany: Koeltz Scientific Books.

Bowker, G. C., & Star, S. L. (1994). Knowledge and infrastructure in international information management: Problems of classification and coding. In L. Bud-Frieman (Ed.), *Information acumen: The understanding and use of knowledge in modern business* (pp. 187–216). London: Routledge.

Bowker, G. C., & Star, S. L. (forthcoming 1999*). Sorting things out: Classification and practice.* Cambridge, MA: MIT Press.

Brandenburg, W. A. (1991). The need for stabilized plant names in agriculture and horticulture. In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 23–31). Konigstein, Germany: Koeltz Scientific Books.

Briquet, J. (1906). *Règles internationales de la nomenclature botanique adoptées par le congrès international de botanique de Vienne 1905 et publiées au nom de la commission de rédaction du congrès (International rules of botanical nomenclature adopted by the Vienna 1905 International Congress of Botany and published in the name of the Congress's commission on naming).* Jena, Germany: Verlag von Gustav Fischer.

Cain, A. J. (1958). Logic and memory in Linnaeus's system of taxonomy. *Proceedings of the Linnaean Society of London, 169,* 144–163.

Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities.* Chicago: University of Chicago Press.

Chauvet, M. (1991). The needs for stability in names for germplasm conservation. In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 33–38). Konigstein, Germany: Koeltz Scientific Books.

Cronquist, A. (1991). Do we know what we are doing? In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 301–311). Konigstein, Germany: Koeltz Scientific Books.

De Candolle, A. (1867). *Lois de la nomenclature botanique (Laws of botanical nomenclature)*. Paris: V. Masson et Fils.

De Smet, W. M. A. (1991). Meeting user needs by an alternative nomenclature. In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 179–181). Konigstein, Germany: Koeltz Scientific Books.

Fifth International Botanical Congress (1931). *Fifth International Botanical Congress, Cambridge, 16–23 August, 1930; Report of proceedings edited for the executive committee by F. T. Brooks and T. F. Chipp*. Cambridge: Cambridge University Press.

Galtier, J. (1986). Taxonomic problems due to preservation: Comparing compression and permineralized taxa. In R. A. Spicer & B. A. Thomas (Eds.), *Systematic and taxonomic approaches in palaeobotany* (pp. 1–16). Systematics Association Special Volume No. 31. Oxford: Clarendon Press.

Graham, L. (1972). *Science and philosophy in the Soviet Union*. New York: Knopf.

Gunn, C. R., Wiersema, J. H., & Kirkbride, J. H., Jr. (1991). Agricultural perspective on stabilizing scientific names of spermatophytes. In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 13–21). Konigstein, Germany: Koeltz Scientific Books.

Hawksworth, D. L. (Ed.). (1991). *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991*. Konigstein, Germany: Koeltz Scientific Books.

Heywood, V. H. (1991). Needs for stability of nomenclature in conservation. In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 53–58). Konigstein, Germany: Koeltz Scientific Books.

Heywood, V. H. (1997). Information needs in biodiversity assessments—from genes to ecosystems. In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 5–20). Konigstein, Germany: Koeltz Scientific Books.

Hughes, N. F. (1991). Improving stability of names: Earth sciences attitudes. In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 39–44). Konigstein, Germany: Koeltz Scientific Books.

ICBN. (1956). *International code of botanical nomenclature adopted by the 8th International Botanical Congress, Paris, July 1954*. Utrecht, Netherlands: IUBS-UNESCO.

IRBN. (1935). *International rules of botanical nomenclature, adopted by the International Botanical Congresses of Vienna, 1905, and Brussels, 1910; Revised by the International Botanical Congress of Cambridge, 1930, compiled by the editorial committee for nomenclature from the report of the subsection of nomenclature prepared by John Briquet* (3rd ed.). Jena, Germany: Verlag von Gustav Fischer.

International Botanical Congress. (1906*). Actes du congrès international de botanique tenu à Vienne (Autriche 1905) (Acts of the International Congress of Botany held in Vienna [Austria 1905]).* R. V. Wettstein & J. Wiesner (Eds.). Jena, Germany: Verlag von Gustav Fischer.

Kirk, P.M., & Cannon, P.F. (1991). Problems facing indexers and the definition of effective publication. In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 279–286). Konigstein, Germany: Koeltz Scientific Books.

Klemm, C. de. (1990). *Wild plant conservation and the law*. International Union for Conservation of Nature and Natural Resources Environmental Policy and Law Paper no. 24. Geneva: IUCN–The World Conservation Union.

Kuhn, T. S. (1957). *The structure of scientific revolutions.* Chicago: University of Chicago.

Little, F. J., Jr. (1964). The need for a uniform system of biological numericlature. *Systematic Zoology, 13,* 191–194.

Lock, J. M. (1991). The index Kewensis and new plant names: Practical considerations. In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 287–290). Konigstein, Germany: Koeltz Scientific Books.

Matthews, R. E. F. (1983). *A critical appraisal of viral taxonomy*. Boca Raton, FL: CRC Press.

Needham, J. G. (1910). Practical nomenclature. *Science, 185,* 295–300.

Nomina Conservanda. (1965). *Nomina Conservanda Proposita II; Proposals on the conservation of generic names submitted to the 11th International Botanical Congress, Seattle–1969*. Utrecht, Netherlands: UNESCO-ICSU-IUBS.

Patterson, D. J., & Larsen, J. (1991). Nomenclatural problems with protists. In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 197–208). Konigstein, Germany: Koeltz Scientific Books.

Poincaré, H. (1905). *Science and hypothesis.* New York: Science Press.

Rich, A. (1978). Cartographies of silence. In A. Rich (Ed.), *The dream of a common language: poems, 1974–1977.* New York: Norton.

Ride, W. D. L. (1991). Justice for the living: A review of bacteriological and zoological initiatives in nomenclature. In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 105–122). Konigstein, Germany: Koeltz Scientific Books.

Stevens, P. F. (1991). George Bentham and the Kew Rule. In D. L. Hawksworth (Ed.), *Improving the stability of names: Needs and options; Proceedings of an international symposium, Kew, 20–23 February, 1991* (pp. 157–168). Konigstein, Germany: Koeltz Scientific Books.

Common Names: Cooperative Access to Databased Natural History Information

Bernadette G. Callery

Abstract

The design and use of electronic information systems to provide cooperative access to natural history museum collections is influenced by existing traditions of organization of and access to paper-based information about those collections. These information systems lack standardized choices for identifying characteristics and the descriptive terms to specify those characteristics. Initially supplied specimen information is often incomplete and inaccurate, as collections are seldom thoroughly described at the time of discovery, and inconsistent, as taxonomic systems change over time. Accession records vary in levels of detail and are frequently supplemented and validated by information scattered through such archival records as field notes, correspondence, photographs, and other visual records. Although these inconsistencies in an individual museum's records could be accommodated by its own staff and the outside users they directly assist, such information systems may not serve the needs of a more diverse group of users, including public and academic users from other disciplines. While the rhetoric of network culture may imply that the technology that enables cooperation will ensure that cooperation, it is not easily achieved, even when endorsed by professional associations and granting agencies. Examples are drawn from the library community to illustrate solutions to this problem of providing electronic access to records of dissimilar form, content, and descriptive vocabularies.

Specific Needs of Museum Information Systems

Museums can be defined as collections of objects assembled and maintained within a specific intellectual environment. In order to maintain that environment, information about the objects in the museum's collections must be available for use by all the audiences served, including museum staff, administrators, regulatory agencies, subject specialists, and the general public. The apparent desirability of electronic access to collection information is changing the focus of public exhibitions and scientific research. Lynch (1998) noted that while curated exhibitions have been the primary means of public access to museum collections, with increased electronic access museums will become more like libraries, where users can impose their own order on the collections. Some mechanism must be in place to capture, maintain, and selectively deliver that information, depending on a user's relative need to know. Sensitive information will need to be masked from certain users but made available to others. This sensitive data includes location of such economically exploitable objects as rare minerals, plants with potentially medicinal value, and rarely found and therefore collectible insects and animals. The information system must accommodate the fact that taxonomic systems used to identify individual specimens are cumulative. Modern nomenclature is based on taxonomic decisions published in the past, sometimes as long ago as the eighteenth century. The scientific name initially assigned to a given specimen may change either because of incomplete or incorrect initial identification or subsequent changes in the nomenclatural hierarchy of the organism.

The accession record, the basic element of a museum's traditional record-keeping system, records the transaction by which the object was acquired and describes the object so that it can be identified throughout its life cycle. During their life-cycle stages, as distinguished by Bearman (1987), museum objects are considered, acquired, accessioned, managed, conserved, documented, studied, interpreted, and deaccessioned or destroyed. The information system must maintain the relationship of the object with information concerning the circumstances of the object's discovery and acquisition and the provenance of past ownership, and allow the accumulation of information about the object—even when it is conflicting—over its lifetime. As nomenclatural information in natural history is notoriously unstable over time, information systems must maintain connections among the original published name of an organism, its currently accepted name, and the historic variants used along the way in taxonomic and bibliographic references. Different levels of specificity, particularly in geographic location information, must also be accommodated by the system, as levels of detail supplied have varied over time and local museum practice.

In paper-based information systems various types of information about the acquisition and identification of a particular specimen were often maintained in multiple files, with long-time museum staff needed to piece together extant records and unwritten museum lore in order to verify questions of circumstances of acquisition or other collection details. As noted in Sarasan's frequently quoted article "Why Museum Computer Projects Fail" (Sarasan, 1981), the inability to incorporate this invisible contextual information into the electronic specimen record was the major cause of failure in early projects to computerize access to specimen information. Acknowledging the need to capture this network of curatorial lore and familiarity with past museum personalities and practices, Sarasan notes that "without oral tradition, many collection information systems would have failed even to fulfill the two basic functions of museum documentation—to lead the user to the specimen in a reasonable period of time, and to interrelate all the information sources so that a user might easily find all the information recorded about a particular object." The institution-specific nature of this information network may argue against cooperative descriptive systems on which the efficiency of shared-access systems is based.

Unlike art objects, which may be cherished for their cultural values, as emblems of power, or even for the raw market value of their materials, natural history museum specimens have neither meaning nor value outside their context of what, where, and when. Physical arrangement of natural history collections carries meaning, as it provides a visual index to the taxonomic context. One proof of the widespread acceptance of this assumption was the simple statement in a popular introductory text to botanical taxonomy that "plants are arranged in the herbarium according to a selected classification" (Lawrence, 1951). This statement introduced the few paragraphs on the relative merits of classification schemes for particular types of herbaria collections and was in marked contrast to the many pages of detailed recommendations for the actual practice of collecting and specimen preparation in the field and in the herbarium.

With the rise of such popular information technologies as the World Wide Web, the public's expectation is increasing that museums will make information on their specimen collections available electronically to a wider range of users than was ever considered possible in the past. Speculating on the role of museums in the electronic age, Sullivan (1998) says that a museum's "walls have become electronically permeable and access to collections in the twenty-first century may become as important as possession of collections was in the twentieth century." To accomplish this in an efficient and timely fashion, museums have been exploring various strategies for providing cooperative access to this electronic information. In the past twenty-five years libraries have benefited from cooperative cataloging initiatives to build their online catalogs, most notably those services provided by OCLC (Online Computer Library Center) and the Research Libraries Group. The museum community, however, has not adopted a similar scheme on the grounds that individual specimens are unique and cannot be managed as uniformly as individual copies of a published book. Lack of a controlled vocabulary, including standardized descriptive terms, limits the effectiveness of searching across collections divided by discipline and institution. Such proponents as Rosenberg (1997) recognize the value of taxonomic databases as guides to both the hierarchical taxonomic structure and standard thesauri (lists of equivalent terms intended to guide the user to the specific vocabulary in a particular information system). He notes that "given the magnitude of the task of capturing data and the paucity of resources for taxonomic pursuits, efficiency in compiling collection databases is critical. Taxonomic databases that document the nomenclature, synonymy, and classification of species and higher taxa can provide greater efficiency and accuracy in computerizing the raw data of collections." But the creation of such cooperative databases has been largely limited to lists of accepted valid names, many of which are available electronically.

Record-Keeping Traditions

The design and use of electronic information systems to provide access to natural history museum collections is influenced by existing traditions of organizing paper-based information about those collections. Record-keeping systems in natural history museums document the work of the institution, which is to collect, to identify, to preserve, and to provide access to the objects in its collection. In these museums the evidential value of the object itself is supplemented, not supplanted, by the documentary evidence of field notes, photographic and other visual records, formal accession information, and published works referring to that specific object. The disciplines of taxonomy and systematics are used to name and relate objects in a museum's collection. They are what distinguishes the modern natural history museum from the Renaissance's cabinet of curiosities.

Museums are part of a long tradition of data recording, analysis, and dissemination. Researchers have recorded data out of sheer curiosity, out of a desire to pass information on to succeeding generations, for self-aggrandizement, and as a show of power. The creation and maintenance of catalogs of collections removes ambiguity. Westbrook (1992) notes that in 150 B.C. Hipparchus of Rhodes created his star catalog, which listed the location and brightness of over a thousand stars, as he was unable to decide whether a given star was really new to science or had simply been inadequately described in the past. Much of the rationale for the record-keeping systems of the Kunstkammern, or cabinets of curiosities, which were accumulated from the mid-sixteenth century through the mid-eighteenth century, is based on much earlier works, including Pliny's *Historia naturalis*, in which the natural curiosities were separated from the man-made ones. Published catalogs of private collections, such as the 1599 *Historia naturale*, which recorded the collection of Ferrante Imperato, or the better-known *Museum Wormianum seu historia rerum rariorum* by Ole Worm, published in Amsterdam in 1655, itemize the holdings of these collections, without much information on the circumstances of collection or records of provenance. Establishing the relationship of one object to another has been the basis of systematics in natural history since the sixteenth-century Moderns declared their superiority over the Greek and Roman Ancients, believing that there was more to nature than there was in Aristotle. Variability in the names of plants and animals, particularly those from exotic locales, was recognized even in fifteenth- and sixteenth-century printed books of flora and fauna. The recording of languages and culture-specific variants of names was an important feature of those publications. One example of the continuing value of this nomenclatural diversity was the inclusion, by reference, of the extensive synonymy of Caspar Bauhin's 1623 *Pinax theatri botanici*, in the major works of the great eighteenth-century systematizer, Carl Linnaeus, including his landmark *Species plantarum*, which is the touchstone for modern botanical nomenclature.

Public Access to Museums

Private museums assumed that the personal attention of the museum's owner would serve as curator and interpret the objects for the individual visitor. As these private collections were institutionalized and made available to the public throughout the nineteenth century, curators attempted to serve as personal guides for the public. These attempts met with limited success, often because of the lack of a background common to both visitor and guide or sheer ineptitude on the part of the guides. Early public museums were criticized for their lack of apparent organization of collections in the public exhibition galleries. A visitor to the British Museum in 1786 noted that except for "some fishes in a small apartment which are begun to be classed, nothing is in order, everything is out of place, and this assemblage appears rather an immense magazine, in which things have been thrown at random, than a scientific collection, destined to instruct and honour a great nation" (Ripley, 1969). The modern separation of specimen collections organized for scientific use from objects selected for public display dates to the mid-1860s when these distinctions were debated by John Edward Gray and Richard Owen of the Natural History Departments in the British Museum and applied by William Flower at the Hunterian Museum (Stearn, 1981). Twentieth-century natural history museums are the product of years of refinement of the concept of the "index museum," where selected typical specimens summarize the whole in a relatively small space. It was felt that collections organized of unique, but related specimens, would bewilder and tire the public.

George Browne Goode, assistant secretary of the Smithsonian Institution in the late 1800s, accepted his museum's place in the public tradition of arrangement and description. He wrote in 1895 that a museum should be "much more than a house full of specimens in glass cases. It should be a house full of ideas, arranged with the strictest attention to system." He then continues wryly that "an efficient educational museum may be described as a collection of instructive labels, each illustrated by a well-selected specimen" (Goode, 1895). The arrangement of the collections from the U.S. Exploring Expedition of 1838–42, or U.S. Ex. Ex. as it was referred to, from the name of the expedition as it appeared on the specimen labels, was one of his responsibilities. The collections amassed during the expedition became the core of the Smithsonian's collections, although there were considerable losses from improper preservation and documentation techniques. When the expedition returned in 1842, no museum was equipped to receive or systematically maintain the artifacts that had been collected. The deposition of the collections in the Smithsonian Institution in 1858, after much discussion and resistance, was one of the major accomplishments of the expedition.

During the third quarter of the nineteenth century in the United States the identification of specimens received from such government-sponsored explorations as the U.S. Ex. Ex. was a major stimulus to the study of natural science. Goode believed that understanding historical processes was essential to describing the present state of scientific knowledge. This approach was consistent with the overall activity in acquiring and organizing documentary evidence, which coincided with the contemporary movement in historical studies led by Herbert Baxter Adams. Concurrent with the movement to collect artifacts of the past was the publication of metaphoric "cabinets," including serial publications that reproduced literary or art works of the past, forming a museum in print. In a more abstract sense specimen collections served as further evidence of the upward spiral of progress and improvement. The cultural historian Henry Shapiro (1985) sees the principal significance of these specimen collections as reminders of "the pastness of the past and as artifacts, of the development of evolutionary sequence that was history . . . and it was as a monument to the distance between past and present, hence proof of the reality of progress and evidence of the character of progress that they were preserved." Throughout the twentieth century, collections continued to grow, as did the costs of housing, preservation, and access. The cost and intellectual burden of rapidly enlarging collections is based on the Darwinian approach that extensive collections are required to elucidate the evolutionary process (McAlpine, 1986). The need for continued growth of specimen collections through additions of "multiple copies" of individual species has been disputed, but generally successfully countered with the argument that a single specimen might be anomalous (Bryant, 1983).

The Rise of Automated Museum Record-Keeping Systems

Sarasan and Neuner's 1983 survey of the computerization of museum collections discusses the collection information crisis of the 1970s. There was an increasing sense of the public accountability to maintain and provide access to objects maintained by museums as a public trust, combined with an increasing acknowledgment of the difficulty of locating specimens and maintaining access to the information associated with them. Responding to this demand for public accountability, the Commission on Museums for a New Century reviewed existing museum practice and in 1984 proposed a number of objectives with corresponding recommendations for achieving those goals. One goal was the use of information management technology as a means of capturing and preserving specimen or item-level collection data. The commission's report defined the ethical and legal obligations of museums "to maintain and manage the objects entrusted to them and that involves all the activities necessary to preserve objects in perpetuity, to gain intellectual control over them (by acquiring and recording information about them) and to make them accessible to scholars" (Commission on Museums for a New Century, 1984).

The development of automated collections management systems was heralded as the solution to internal and external demands for access. Several museum-developed software packages had appeared in the 1960s, including the Smithsonian's Self-Generating Master (SELGEM), succeeded by the Collection Information System (CIS), an integrated collections management system that combined ongoing specimen documentation of an object's movement through the phases of exhibition, conservation, and loan in a single system. In her survey of North American collection management systems, Sledge (1988) believed that the rush to develop and install such systems was driven by the expectation that automated audit control of collections could be performed if a collection inventory existed. What resulted was a separation of curatorial staff from system developers, even though some curators expanded their traditional responsibilities to include the ability to specify and design an information system. Collection administrators, however, are ambivalent about the merits of this change in curatorial responsibilities, as noted in a recent Association of Systematics Collection (ASC) publication (Zorich & Hoagland, 1995) in which curators are warned against "losing" staff to database management. Sledge's comments that the major effect of the incorporation of computers in a record-keeping system was to highlight the inconsistencies of the existing manual systems were repeated throughout the museum documentation literature.

The British Museum Documentation Association (MDA) was established in 1977 to "assist museums with documentation procedures." The MDA published *Practical Museum Documentation* in 1980, with a revised edition in 1981. This work was intended as a guide to techniques for documenting a museum collection. At the time of publication the editors were unaware of any similar publication, although they did note several recent books on museum registration techniques, including Orna and Pettit's 1980 *Information Handling in Museums* and Dudley and Wilkinson's 1970 *Museum Registration Methods.* Features of an effective museum documentation system, capable of satisfying the "demands of a user, whether a curator, researcher or member of the public," were described in the 1981 edition of *Practical Museum Documentation*. In addition to being able to handle any number of records of varying length, the ideal museum documentation system should be easy to use and provide quick access to information but have the ability to block general access to whatever types of information are considered confidential.

In the late 1980s the number of publications dealing with collections management systems for museums increased. Collections management had became the fashionable phrase in the 1980s as museums refined their functions and focused more on the need for effective care of their collections than on acquisition. Collections management was the theme of the MDA's first annual conference in 1987, and its collected papers were prefaced with the statement that museums were working to "control collections and demonstrate accountability, . . . as new computer systems became available for local adoption" (Roberts, 1988). Collections managers were becoming aware of the opportunities for intellectual access as an outgrowth of inventory control and were beginning to see the new information technologies as a way to provide information for research and management at a number of different levels, without multiplying the actual number of records associated with an individual object. The museum accession record was being transformed into the integrated specimen data record. Spiess (1988), whose paper in that collection addressed the policies and procedures of the Smithsonian Institution, stated the need to provide museum-wide access to collections and their associated data in order to support both public education and research and provide an efficient tool for collection management.

The struggle for standards is most evident in the literature dealing with the transition from a manual to an automated record-keeping system. In the introduction to the 1988 edition of Chenhall and Vance's *Museum Collections and Computers* the authors compare this work to their 1975 *Museum Cataloging in the Computer Age*, noting that while the basic principles of museum cataloging had not changed, the power of the indexing and access tools available to individuals had. The similarity between museum accession records and library catalog records is stressed, perhaps hoping to console the museum administrators, who formed part of the stated audience of the book. Since libraries had clearly solved the problem of automated collection management, it was assumed that some of their experiences might benefit the museum community. But the contrasts were found to be too great, and the authors recommend the development of separate information systems unique to each museum so that the special needs and requirements of the individual scientific disciplines and the museum administrations themselves might be best accommodated. Chenhall and Vance (1988) believed that it was "not feasible to develop one 'ideal' cataloging system that will adequately serve a large number of museums, and in the process, allow the free and easy electronic interchange of all data about all objects in all the museums. The Canadian Heritage Information Network has demonstrated that a single system can serve an entire country, but even with this system it is still necessary for each institution to determine the information that it needs or wants to put into the system." This work also provides a substantial body of information on what the authors acknowledge as a "rather esoteric field of specialization" dealing with the standardization of interpreted erratic manual practices of the museum accession file, particularly when multiple data files are linked to an acquisition record. While this work was meant as an introduction to the capabilities of databases and the jargon of networking, its bibliography serves as an overview of the museum computer resources and organizations as well as the literature available at the time.

Response from the Professional Organizations

The ASC, founded in 1973, included a Council on Standards for Systematics Collections, which, in addition to developing standards for specimen and data acquisition and documentation, also recognized the importance of electronic data processing for recording and retrieving specimen information. Much of the current work being done in the systematics community is supported by the ASC and deals with the construction of data models and controlled vocabulary in the area of locality data, especially in stabilizing variant forms of place names. As

Bearman (1989) wrote, "terminology standards are the finest sieves in the hierarchy of information standards." The ASC has issued position papers recommending policies for sharing and use of electronic specimen data (Hathway & Hoagland, 1993; Hoagland, 1994). These guidelines for institutional policies and planning emphasize the significance of accession information as the primary record of accountability. "Accessions result in tangible assets that are held in public trust. Accession policies establish the legal and ethical basis for acquisition and ownership of collections. They are the basis for establishing institutional control over specimens" (Hoagland, 1994).

The British MDA was involved in the development of a standardized system of recording for museums known as the Museum Documentation System, which was adopted throughout the United Kingdom. The MDA proposed several Database Management Systems (DBMS) oriented for specific functions of museums (Thompson, 1992). During this early database period the core set of required fields was hotly debated, with the upshot being that no two museums, even within the same discipline, could wholly agree.

An examination of the professional literature dealing with the documentation of natural history collections—particularly as museums prepare to automate these systems as a means to providing electronic access—shows a struggle toward the adoption of data standards, both in the types of information collected as well as the terms used to describe that information (Moritz, 1989). Members of the individual scientific disciplines have taken responsibility for the definition of the actual fields of data to be collected, producing thesauri and authority files of valid genus and species names, and a wide range of other documentation standards. Effective thesauri depend on authority files, that is, lists of accepted terms for the names of organisms, subject terms, or other descriptors. Use of terms other than those in the authority files, particularly for indexing purposes, results in scattering of references and loss of potentially relevant references to the user.

While the Getty Art History Information Program (AHIP), now the Getty Information Institute, has taken the lead in developing standards for the electronic interchange of images and textual data related to collections held by art and cultural heritage museums, there is no similar widely acknowledged leadership in the development of standards for descriptions of natural history collections. One hypothesis for this delay is that museum curators consider each item in the collection unique, so that attempts at cooperative cataloging have little value for the individual museum except as an academic exercise in cooperation.

Cooperative access models in the art museum and library community are closer to the union catalog approach of bibliographic databases, sharing a belief in common descriptive practices, including the consistent use of controlled vocabulary, as the basis for cooperative cataloging and access. However, Bower (1993), writing from the vantage point of the Getty AHIP, argues against the combination of individually created databases into a single resource because errors of ambiguity may be introduced. "Data that are unambiguous within the context of their initial capture . . . may become ambiguous when juxtaposed with data in different languages, data from other disciplines where overlapping terms have not been rendered referentially unique, or data from the same discipline that use different but equivalent terms to express names and concepts."

And what of the solutions proposed by the library community for addressing the problem of integrating specimen data into museum records? Several recommendations were made that the MARC (Machine Readable Cataloging) format be used to describe museum objects, particularly given the development of the Archives and Manuscript Collections format and specific visual materials formats. Bierbaum (1990) suggests that museums and libraries are alike in creating surrogate records for objects in their collections but cautions that converting museum records to MARC may not go smoothly because of the lack of descriptive standards in the museum community. Reporting on the use of the MARC structure at Berkeley to provide access to a range of nonbook collections, Besser and Snow (1990) present the mutually exclusive options of a specifically designed relational database and the existing MARC standard. They remind us that for every opportunity there is a corresponding obligation, noting that "the flexibility that one enjoys in a relational database management system avoids the stricture of the MARC structure that the bibliographic retrieval systems require, but one pays for that in lack of consistency and transportability." Bearman (1989, 1990) speaks of the advantages of shared reference files and the development of generally useful thesauri, even though a union catalog of (almost) unique items does not have the economies of scale that a union catalog of print-based materials would have. To date, there have been a few experiments with using the MARC format as a vehicle for describing museum specimens, but they are considered novelties.

Proposed Data Models for Cooperative Access to Natural History Museum Information

While the rhetoric of network culture may imply that technology that enables cooperation will ensure that cooperation, such cooperation is not readily achieved. Nomenclatural differences, as well as differences in the database schema, that is, the selection of which data are collected and at what level of detail, have tended to balkanize developing biological databases (Williams, 1997). With the popularity of the Web as a mechanism for allowing access to specimen information and other collection data once maintained solely on institutional databases, more efficient handling of data in distributed repositories has become a major issue (Schatz, 1997). Proposed solutions include "automatic" generation of hyperlinks among "federated" databases (Jamison, Mills, & Schatz, 1996) as a means of relating conflicting names. Data discovery techniques used in data mining may also be used to help develop algorithms that enable the capture and use of historical data.

As expectation for access to these scattered data resources grows, such techniques as vocabulary switching have been proposed as a method to assure interoperability between nomenclatural systems. In bibliographic practice see and see also references are a type of vocabulary switching. This technique is used to preserve relationships between terms and make a user aware of additional subject terms or alternate forms of a name used in a catalog that they would have otherwise missed. Application of vocabulary switching techniques as discussed by Tillett (1991) could take advantage of existing relationships between alternate scientific names and allow the retention of older or inconsistent nomenclature.

Directories of Electronic Resources

One potentially valuable cooperative development along these lines is the recently announced project to develop directories of taxonomists and natural history collections available via the Internet. The National Biological Service has signed an agreement with the ASC to develop these directories, which will be available through the National Biological Information Infrastructure (NBII). The collection survey will include information on the status of collection information automation projects and hot links to those databases, when permitted. Given the proliferation of electronic records forming the publication base for such projects as the Flora North America project and other floristic projects arising from the research efforts coordinated by the Missouri Botanical Garden, an undeniable need exists for a similar system of pointers to the location of this material. There is also a growing need for the documentation of collaborative projects, such as NATUREnet, wherein nine of the large U.S. natural science museums and two botanical gardens are exploring the possibilities of shared specimen records.

While adherence to a model similar to that of the bibliographic standards employed by libraries has been suggested, the systematics community has not adopted it. The union listconcept of specific item-level holdings (e.g., of particular issues of a journal) to which individual libraries contribute according to mutually accepted standards does not have any serious followers. Instead the individual disciplines of natural history maintain their own specimen level information, usually on a per-museum basis. Information sharing consists of allowing access to individual searchable files maintained by a single museum or by periodically contributing information to an established discipline-based database. Since some manipulation of the contributed data is required to make it conform to the depository database, interactive updates of data are not easily done.

Community Standards

The publications of the ASC, particularly the *ASC Newsletter*, are perhaps the best indicator of the progress of the systematics community in arriving at a series of standards for description of specimen collections. The expectation that a standard for data exchange, even within a single scientific discipline, would emerge from endless working groups, such extensive self-studies as the MITRE report (Cooley, Harrington, & Lawrence, 1993), and admonitions from such theorists in the museum computer field as David Bearman has not materialized. In the August 1995 issue of the *ASC Newsletter*, which reported on a July 1995 symposium during the organization's annual meeting titled "Natural History Collections on the Information Superhighway," members were "challenged . . . to accept the concepts of shared databases and centralized software system development." If accepted, this concept would represent a major shift in the systematics community's approach to access to electronic specimen information. The development of "middleware" that would allow for the long-desired interoperability among autonomous systems continues to be discussed, with more systematists involved in the development of data models and controlled vocabulary lists. Other major concerns of the ASC computer and networking committee included the investigation of the ability to publish information from combined databases and the development of shared authority files.

During the 1995 ASC annual meeting Stanwyn Shetler from the National Museum of Natural History, Smithsonian Institution, spoke on ASC's strategic planning goals. "The time is ripe for museums to usher in the Age of Access. If our generation doesn't figure out how to provide better access to the information stored on our existing collections, then the next generation may not be able to defend keeping these collections. This is a time to consolidate collections, electronically if not physically, and concentrate on improving access to their information, while focusing our future collecting efforts to address specific questions. We continue to amass collections faster than we can assimilate, curate and study them—to store far more information than we can retrieve—all on the assumption that we must fill the museums while we still can, so that some future generations, after everything is extinct, can sit in a sterile laboratory and study the specimens we never got to" (Shetler, 1995).

Behind the ongoing debate about cooperative access to specimen data is a growing concern about the retention of the specimens themselves. Arguments for "pulping the herbaria" and otherwise discarding specimen collections after recording the collections in some digitized form are countered by museum curators who persist in their belief that the physical object, along with its related literature, remains the validating evidence. The ASC's 1993 publication, *ASC Guidelines for Institutional Database Policies* (Hathway & Hoagland), specifically recommend that specimen collections, with their associated documentation, be retained even after the information about the specimen is captured in a systematic database or included in a published monograph. The principal argument for the retention of both specimens and their associated documentation is that errors in understanding and interpretation do occur. Specimens and their original and accumulated documentation should be retained as a means of resolving later conflicts of opinion.

Gateways and Cooperatives

While it is often the case that museum departments function with considerable autonomy within a given museum, curatorial staff are aware of the benefits of maintaining electronic data on collections consistent with developing standards in the museum community. There is value perceived in being conversant, if not necessarily compliant, with standards and practices in the intellectual environment outside the institution. Agencies providing major grant funding, such as the National Science Foundation, increasingly require the ability to share electronic data of a museum's systematic collections as a prerequisite for consideration for further funding.

In 1996 the White House Subcommittee on Biodiversity and Ecosystem Dynamics, in recognizing the value of specimen collections, "identified systematics as a research priority that is fundamental to ecosystem management and biodiversity conservation" (Waggoner, 1997). The Integrated Taxonomic Information System (ITIS) is designed to support improvements in the organization of and access to standardized nomenclature. The success of ITIS, a component of the National Biological Information Infrastructure (http://www.nbs.gov/nbii), depends on the willingness of systematists to contribute taxonomic data on the biota of North America to the database. Systematists can support the NBII's primary objective, which is to increase access to distributed sources of biological data and information, by contributing metadata that describe their electronic collection data to the NBII clearinghouse. Users will search the NBII Metadata Clearinghouse to locate biological data from a distributed network of cooperating information sources and provide links to those individual sources.

The gateway approach used by the NBII initiative is similar to cooperative efforts rising from the academic and museum community. Given the difficulties discussed above in the area of inconsistent nomenclature and descriptive standards, most projects focus on identifying resources and providing access to diverse collections within a single institution. The Berkeley Museum Informatics project, officially begun in 1992, but with considerable preliminary activity dating back to 1987, was created to work with "faculty, collections managers, and curators to develop data models, system architectures and demonstration and production systems as bases for coordinated and integrated approaches to the application of information technology in museums and archives"(http://www.mip.berkeley.edu/mip). Through creating standards for shared access to accession records and catalog information, the project seeks to identify previously isolated collections and their associated information and make this information known and available to a multidisciplinary community.

The Biodiversity and Biological Collections Web server at the University of Kansas (http://biodiversity.uno.edu), originally the Biodiversity and Biological Collections Gopher at Cornell University, also serves as a clearinghouse, identifying searchable resources of interest to systematists. Its MUSE project and the associated workshops in the early 1990s did much to build awareness of the opportunities to cooperate on the computerization of natural history collections.

Conclusion

Natural history museums serve the basic human need to collect, combined with the related goal of science to organize information. Individual scientific disciplines have developed separate organizational schemes with diverse nomenclatural and descriptive structures. Even given present technologies, inconsistencies cannot be resolved at the level of the individual specimen record, particularly in legacy data that are incomplete by contemporary standards. Current solutions to the access problem include creating clearinghouses that identify the existence of information on a particular topic, continuing discussions toward the development of multidisciplinary standards that specify core or essential data in a record, and agreeing on the use of metadata to define and describe the nature and content of information contained in these data sources. A valuable lesson to be learned from the experience of cooperative access to databased information is that every cooperative opportunity has the corresponding obligation to develop and maintain standards if the goal of interactive data interchange is to be achieved. In looking forward, we should also look back to the Smithsonian's G. Brown Goode (1895) who reminds us that "catalogs are the keys to the treasure-vaults of a museum."

References

Bearman, D. (1990). Can MARC accomodate archives and museums? Technical and political challenges. In T. Petersen & P. Moholt (Eds.), *Beyond the book: Extending MARC for subject access* (pp. 237–245). Boston: G. K. Hall.

Bearman, D. (1989). A framework for museum standards. *Spectra, 16*(2), 1–5.

Bearman, D. (1987). Functional requirements for collections management systems. *Archival Informatics Technical Report, 3,* 1–87.

Besser, H., & Snow, M. (1990). Access to diverse collections in university settings: The Berkeley dilemma. In T. Petersen & P. Moholt (Eds.), *Beyond the book: Extending MARC for subject access* (pp. 203–224). Boston: G. K. Hall.

Bierbaum, E. G. (1990). MARC in museums: Applicability of the revised visual materials format. *Information Technology and Libraries*, *9*(4), 291–299.

Bower, J. M. (1993). Vocabulary control and the virtual database. *Knowledge Organization, 20*(1), 4–7, 21.

Bryant, J. M. (1983). Biological collections: Legacy or liability? *Curator, 26*, 203-218.

Chenhall, R. G., & Vance, D. (1988). *Museum collections and today's computers.* New York: Greenwood Press.

Commission on Museums for a New Century. (1984). *Museums for a new century: Report of the Commission on Museums for a New Century.* Washington, DC: American Association of Museums.

Cooley, G. P., Harrington, M. P., & Lawrence, L. M. (1993*). Analysis and recommendations for scientific computing and collections information management of free-standing museums of natural history and botanical gardens.* McLean, VA: MITRE.

Goode, G. B. (1895). *The principles of museum administration*. New York: Coultas & Volans.

Hathway, E. C., & Hoagland, K. E. (1993). *ASC guidelines for institutional database policies.* Washington, DC: Association of Systematics Collections.

Hoagland, K. E. (1994). *Guidelines for institutional policies and planning in natural history collections.* Washington, DC: Association of Systematics Collections.

Jamison, D. C., Mills, B., & Schatz, B. (1996). An extensible network query unification system for biological databases. *Computer Applications in the Biosciences, 12*(2), 145–150.

Lawrence, G. H. L. (1951). *Taxonomy of vascular plants*. New York: Macmillan.

Lynch, C. (1998). Goals of the Coalition for Networked Information (CNI), presentation at the University of Pittsburgh, School of Information Science, 23 June 1998.

McAlpine, D. F. (1986). Curators and natural history collections: Have we become islands in science? In *Proceedings of the 1985 Workshop on Care and Maintenance of Natural History Collections* (pp. 7–14). Toronto: Royal Ontario Museum.

Moritz, T. (1989). Toward community standards in systematics. *Spectra, 16*(2), 18–20.

Museum Documentation Association. (1981). *Practical museum documentation*. Duxford, Cambridgeshire, England: Museum Documentation Association.

Ripley, D. (1969). *The sacred grove: Essays on museums*. New York: Simon and Schuster.

Roberts, D. A. (Ed.). (1988). *Collections management for museums*. Cambridge: Museum Documentation Association.

Rosenberg, G. (1997). Biotic databases, the interface of taxonomic and collection databases. *ASC Newsletter, 25*(2), 13–14, 21.

Sarasan, L. (1981). Why museum computer projects fail. *Museum News, 59*(4), 40–49.

Sarasan, L., & Neuner, A. M. (Comps.). (1983*). Museum collections and computers: Report of an ASC survey.* Lawrence, KS: Association of Systematics Collections.

Schatz, B. R. (1997). Information retrieval in digital libraries: Bringing search to the Net. *Science, 275*, 327–334.

Shapiro, H. (1985). Putting the past under glass: Preservation and the idea of history in the mid-nineteenth century. *Prospects: An Annual of American Cultural Studies, 10*, 248–278.

Shetler, S. G. (1995). ASC strategic planning: Diverse institutions, common goals. *ASC Newsletter, 23*(4), 49.

Sledge, J. E. (1988). Survey of North American collections management systems and practice. In D. A. Roberts (Ed.), *Collections management for museums* (pp. 9–17). Cambridge: Museum Documentation Association.

Spiess, K. P. (1988). Collections management policy and procedures: Initiatives at the Smithsonian Museum National Museum of American History. In D. A. Roberts (Ed.), *Collections management for* *museums* (pp. 133–144). Cambridge: Museum Documentation Association.

Stearn, W. T. (1981). *The natural history museum at South Kensington*. London: Heinemann.

Sullivan, R. (1998). The object in question: Museums caught in the net. In M. T. Wolf, P. Ensor, & M. A. Thomas (Eds.), *Information imagineering: Meeting at the interface* (pp. 147–154). Chicago: American Library Association.

Thompson, J. M. A. (Ed.). (1984). *Manual of curatorship: A guide to museum practice*. London: Butterworth-Heinemann.

Tillett, B. B. (1991). A summary of the treatment of bibliographic relationships in cataloging rules: Linking devices from Panizzi to AACR2. *Library Resources and Technical Services*, *35*, 150–158.

Waggoner, G. S. (1997). The Integrated Taxonomic Information System (ITIS) and Taxonomic Resources & Expertise Directory (TRED): Two online taxonomy-related databases. *ASC Newsletter, 25*(5), 59.

Westbrook, J. H. (1992). A history of data recording, analysis and dissemination. In P. Glaeser (Ed.), *Data for discovery: Proceedings of the 12th International CODATA Conference* (pp. 430–460). New York: Begell House.

Williams, N. (1997). How to get databases talking the same language. *Science, 275*, 301–302.

Zorich, D., & Hoagland, K. E. (1995). *Status, resources and needs of systematics collections*. Washington, DC: Association of Systematics Collections.

The Evolution of the Secondary Literature in Chemistry

Helen Schofield

Abstract

Information scientists and librarians commonly define secondary publications as tools that facilitate identification of relevant primary publications. In this paper, the early history of secondary publications in general is considered briefly, commencing with the first abstracting journal, the *Journal des Sçavans* in 1665. The paper then concentrates on the evolution of the secondary literature in chemistry. Chemistry was one of the first branches of knowledge for which specialist secondary sources were published, the first such publication being Crell's *Chemisches Journal für die Freunde der Naturlehre* in 1778. The reasons for the need for secondary sources in chemistry are examined. Histories of several of the most important secondary sources in chemistry are given, including, in some cases, short biographies of their originators. The differences in editorial policy and criteria for inclusion of publications among the secondary sources are discussed. The organization of information is presented along with how classification schemes were developed to aid retrieval of data and references, since most chemical information requests are based on the need for information about chemical compounds. The evolution of index systems in the pre-computer era, which make use of chemical names and formulae to enable access to chemical information, is described. The ability of systems to deal with new types of substances as they have been discovered is also discussed. A list of major secondary sources in chemistry, both those still published and those that have ceased publication, is given at the end of this article.

Introduction

Information scientists and librarians commonly define secondary publications as tools that facilitate identification of relevant primary publications (Neufeld & Cornog, 1983). They usually provide bibliographic references and brief descriptions as well as subject terms and indexes that can be used to locate many types of primary information sources, such as journal articles, reports, patents, and conference papers. Nowadays, the most common product of such an exercise is a serial publication, either printed or electronic, that covers the literature produced in a specific period.

Every researcher is aware that finding papers on a specific subject can be problematic because of the vast number of journals and other primary sources available. *Chemical Abstracts* today covers approximately nine thousand journals from ninety-three countries in fifty languages, not to mention the patents, conference papers, and other primary sources covered, leading to the publication of about half a million abstracts per year. But when did the volume of literature generated through scholarly research become unmanageable, and how were secondary sources such as *Chemical Abstracts* conceived?

The object of this paper is to consider the origins of the major abstract publications and other secondary literature in chemistry, that is, the publications whose main aim is to organize and index the primary publications, the tools that chemists use or have used to identify primary references. This paper does not consider encyclopedias, monographs, reviews, or works that are purely data compilations and stops short of the computer era. Also excluded are modern publications specifically for current awareness. This paper concentrates on British, American, and German sources, which have traditionally dominated the secondary literature in chemistry.

The Need for Secondary Sources

In the seventeenth and eighteenth centuries the emerging format for learned publication was the journal article. In that period the number of journal titles grew from about thirty-five to about four hundred journal titles. By 1900 there were about 5,000 scientific journal titles being published and 136 abstract journals to cover the articles in these journals (Kaser, 1995). By 1950 there were about 30,000 primary scientific journals. Chemistry was one of the first branches of science to publish abstracts (Crosland, 1994). By the beginning of the twentieth century the pressure on the scientist to keep up to date with the literature was becoming severe and occupied an ever-increasing proportion of the day (Williams, 1977). Thus, abstract journals and other secondary publications came into being. They summarized the material in primary publications to indicate the content of the original papers and acted as a tool for tracing relevant references. From them the reader could judge whether it was worthwhile reading the original article.

Origins of Abstracts

The collection and organization of knowledge can be traced back five thousand years to clay tablets and papyrus scrolls (Skolnik, 1982). The invention of the printing press around 1450 heralded the beginning of the information age, enabling information to be published, stored, and disseminated as never before. Abstracts originated in the Middle Ages; the word abstract is derived from the Latin *abstractus*, meaning "to draw away." Monks would often write marginalia summarizing documents they were transcribing. Kings often required their generals and ambassadors to write summaries of their reports, and since the eleventh century, the Vatican has abstracted reports from its envoys.

On 5 January 1665 Denis de Sallo issued the first number of the first published abstract journal, *Journal de Sçavans*, which was published weekly (Cooper, 1982; Collison, 1971a, 1971b). The journal was part reviews and part abstracts, and each item occupied about half a page. The summaries or reviews covered books, decrees, or informative letters, with the primary publication's author, title, and place of publication. De Sallo can therefore be considered the inventor of the abstract journal, although he was only personally involved for the first thirteen issues. The journal continued until 1792.

Other early abstract publications include the *Nouvelles de la République des Lettres* (1684–1718) and *Histoire des Ouvrages des Savans* (1687–1706; 1708–1709), both published by French people living in exile. The first German abstract publication was *Monatsextracte*, which commenced in 1703. The famous *Aufrichte*, which abstracted about forty journals, was published between 1714 and 1717. The first abstract journals published in England were the *Universal Magazine of Knowledge and Pleasure* (1747–1815) and the *Monthly Review* (1749–1844).

Origins of Abstracts in Chemistry

Abstracts in Primary Journals

It became common for primary scientific journals to publish abstracts of work reported elsewhere in addition to original papers, often in separate sections titled "News from the Literature" or something similar. The *Philosophical Transactions of the Royal Society*, for example, published abstracts. The first such publication in chemistry was *Crell's Chemische Journal für die Freunde der Naturlehre* (1778–81). This was the first in a series of journals published by Lorenz Florenz Friedrich von Crell (1744–1816), a professor at Braunschweig and then Helmstadt and Göttingen. Subsequent publications were *Chemisches Annalen für die Naturlehre, Arzneygelahrtheit, Haushaltungskunst und Manufacturen* (1784–1803), *Beiträge zu den chemischen Annalen* (1785–99), and *Neueschemisches Archiv* (1784–91). These provided a forum for German chemists to exchange their views and aided dissemination of information. They became models for publications in Germany and elsewhere. A number of French chemistry journals published abstracts (Crosland, 1994), including *Bulletin de la Société Chimique de France* (from 1863), *Annales de Chimie* (from 1851), and *Comptes Rendus*, the last composed mainly of abstracts because of its inability to attract original work. The Deutsche Chemische Gesellschaft published abstracts in its *Berichte* from 1868 to 1896. (In 1896 it assumed responsibility for publishing *Chemisches Zentralblatt*.) In Britain abstracts were published in the *Journal of the Chemical Society*, beginning in 1871, and *Journal of the Society of Chemical Industry* beginning in 1882 (Whiffen, 1991). Another British journal to publish abstracts was *The Analyst*. A number of current primary journals still publish abstracts in this way.

Early Secondary Chemistry Sources in Germany and France

Chemistry was one of the first subjects to have secondary publications. Berzelius was prompted by the increasing amount of journal literature to begin *Jahresberichte über die Forschritte der physichen Wissenschaften* in 1829, the first review journal that concentrated on chemistry, continuing for over twenty-five years. According to Crosland (1994), abstracts were first published in France in 1858 by the Société Chimique de France in the *Répertoire de Chimie Pure*. Crosland quotes Charles Adolphe Wurtz, the editor of *Répertoire*, "This journal is intended to put the public of our country in touch with the progress of pure chemistry in France and abroad. A publication of this kind does not yet exist in our scientific literature and seems to meet a real need. It will accept no original work but will offer to the reader a summary [of what has been published elsewhere]." There was also the *Répertoire de Chimie Appliquée*, edited by the French industrial chemist Charles Louis Barreswil, covering applied chemistry (Manzer, 1977). The *Répertoires* were cover-to-cover abstract journals, containing unnumbered abstracts organized into subject categories with annual author and subject indexes. They were replaced by the *Bulletin de la Société Chimique de France*, which in 1863 started to publish abstracts.

Funding and Sponsorship

Funding and sponsorship of secondary publications came from personal sponsors, learned and professional societies, industrial research institutes, government agencies, and commercial enterprises (Cooper, 1982). In the nineteenth century the Société Chimique de France, Deutsche Chemische Gesellschaft, Chemical Society, and Society of the Chemical Industry were important sponsors. A strong chemical, and later pharmaceutical, industry had created extraordinary demand for chemical information. Chemistry's industrial links have always ensured that funding is available for research into the organization and dissemination of chemical information. Between 1900 and 1945 there was exceptional growth in the chemical industry, an increase in the number of chemists and engineers, and a corresponding growth in the literature. Nevertheless, secondary sources were not guaranteed to be commercial successes (Whiffen, 1991).

The Language of Chemistry

According to Bowman (1974), most requests for chemical information focus on chemical compounds, with chemists commonly asking questions of the following types: What are the properties of this compound? How can I make this compound? What compounds have the following properties? What compounds similar to this one exist? It has been estimated that approximately 85 percent of index entries in the 1966 subject index to *Chemical Abstracts* are associated with compounds and materials (Tate, 1967; Whittingham, Wetsel, & Morgan, 1966). Information retrieval systems must be able to provide answers to these questions. The early producers of secondary information sources in chemistry recognized the central role of the chemical compound and developed systems accordingly.

Unlike some other scientists and technologists, chemists often consult older literature. Tate (1967) suggested that it was not uncommon for chemists to go back sixty years and find immediately useful information. This is borne out by the present author's experience of working in a chemistry library, where material up to a hundred years old can still be of use, particularly to organic chemists. The information required, as stated above, is usually concerned with chemical structures, hence the importance of understanding older chemical indexing and nomenclature systems even today.

The development of the language of chemistry is beyond the scope of this paper and is well documented (Crosland, 1962). Communication of chemical substance information depended initially on trivial names before systematic nomenclature schemes were devised. Use of trivial names and different systematic nomenclature systems can create problems. Chemical names vary between languages; for example, many early secondary sources were in German, a problem for non-German speakers. Other problems arise with complicated structures. Although the publishers of the secondary sources employ nomenclature experts who standardize chemical names, few practicing chemists have sufficient knowledge of the nomenclature schemes used by different sources to allow derivation of accurate names. This disjunct between designers and users has commonly led to the production of empirical formula indexes, although these do not solve the problem completely, as often many structures can be drawn from one empirical formula.

Computers solve many of these problems by enabling the two- or three-dimensional structure to be the means of communication. In the pre-computer era some secondary sources overcame these problems in part by use of innovative classification schemes for compounds, which could appear complex but, once mastered, ensured effective retrieval of information. Such schemes were used in the *Beilstein Handbuch der organischen Chemie* and *Gmelin Handbuch der anorganischen Chemie*.

Gmelin's *Handbuch der Anorganischen Chemie*

An early information source is the *Gmelin Handbuch der anorganischen Chemie*, founded by Leopold Gmelin (1788–1853). Some might consider the Gmelin and Beilstein handbooks to be tertiary sources of information, (defined by Mellon [1965] as aids to searching the secondary and primary sources, such as guides to the literature or publications that provide facts about chemists and their work, for example, directories and dictionaries; they are more often used by librarians than practicing chemists), but they are so frequently considered the main sources alongside *Chemical Abstracts* that they are discussed here.

Gmelin's Life

Leopold Gmelin came from a family closely involved with chemistry (Gillespie, 1970–1980). Gmelin graduated in 1804 and then worked in the family apothecary in Tübingen. During the period from 1700 to 1860 (Walden, 1954) about a dozen members of the Gmelin family were professors in the three university cities of Tübingen, Heidelberg, and Göttingen. Apparently three kinds of Gmelin professors were referred to: those who had passed away, those on the lecture platform, and those in the cradle. Walden reproduces a family tree of the chemistry professors, derived from the publication *Stammbaum der Familie Gmelin* published in Karlsruhe in 1877.

Gmelin was awarded his medical doctorate in 1812, but he was also trained as a chemist and had a keen interest in mineralogy and geology. Gmelin was appointed docent at Heidelberg in 1813, became extraordinary professor in 1814, and was appointed director of the Chemical Institute in 1817. Gmelin worked hard to improve the teaching of chemistry, but he also continued his research. He published papers on physiology, inorganic and organic chemistry, mineralogy, and the theory of chemistry.

The Handbuch

The first edition of Gmelin's *Handbuch* was published in three thin volumes between 1817 and 1819, as *Handbuch der theoretischen Chemie* (*Handbook of Pure Chemistry*). In this book Gmelin reviewed all chemistry, organic as well as inorganic (Skolnik, 1982). The book was conceived to assist Gmelin with his lectures, but it was commercially very successful. His aim was "to arrange systematically all the precisely determined facts concerning every element and compound, to state these facts succinctly and accurately, and also give the pertinent references to the literature" (Walden, 1954). He used the term "organic chemistry" for the first time in German textbooks; he spoke of "imponderable elements" (heat, light, and electricity), as well as "ponderable elements" (the forty-eight elements known at the time); and he coined some new terms, such as *ester* and *ketone*. Further editions followed quickly, and the organic part of the second edition was translated into French. By the time of the fourth edition, published between 1843 and 1852, the work had changed its name to *Handbuch der Chemie*. By 1870 there were ten volumes. The handbook was the most important book of chemistry for more than a generation, and it had a remarkable impact on the development of the science. In addition to his systematic organization of the information within the handbook, Gmelin devised his own arrangement of the fifty-one elements known at the time of the third edition, that is, his own periodic table (*Synergisms in Chemical Information*, 1992). The periodic table organized the known elements in a horseshoe arrangement and was ordered according to the affinities of the elements rather than the atomic number order adopted by Mendeleev and Meyer in 1869. The table was revised in the fourth edition and is now almost forgotten.

Gmelin was solely responsible for the first three editions of the handbook. He also edited the first four volumes of the five-volume fourth edition and the fifth of these volumes was prepared by Gmelin's associates, Karl List and Karl Kraut. They continued the publication on a part-time voluntary basis after his death (part way through the production of the fifth edition). A translation of the fourth edition by H. Watts was published by the Cavendish Society between 1848 and 1872, in nineteen volumes (Skolnik, 1982). The fifth edition was concerned only with inorganic chemistry. In all, five editions had been published in fifty years. The sixth commencing publication in 1872 and seventh editions were edited by Kraut, and the work became known as *Gmelin-Kraut* for that period. In 1921 publication was taken over by the Deutsche Chemische Gesellschaft.

Erich Pietsch served an important role in the progress of the *Handbuch* in the twentieth century (Oesper, 1949). Pietsch was still at the university in Berlin when he was appointed to the editorial staff of the handbook as a part-time assistant. In 1927 he was promoted to an assistant editorship and head of section. In 1935, the Deutsche Chemische Gesellschaft decided to enlarge the staff, and Pietsch was chosen to work out the plan for the expansion and to implement it. He became the head of the Gmelin Institute on 1 January 1936 and continued throughout World War II. After the war conditions were not favorable for production of the handbook, but the British and American governments gave their support as its importance was realized.

In 1946, the Gmelin Institute was placed under the Max Planck Society for the Advancement of Science, in conjunction with the Deutsche Chemische Gesellschaft. The eighth edition is the latest, and in it, Pietsch expanded the scope of the work. The text of the eighth edition was compiled without reference to earlier editions: Each topic is covered by reference to the original sources. By 1948 there were about sixty scientists on the editorial staff and a similar number of technical, presumably production, staff. Pietsch introduced information about ferrous metallurgy, partly because he perceived a gap in the literature, and he added the *Gmelin Patentsammlungen*, coverage that is particularly important for metallurgy. It was prepared in collaboration with the Reichspatentamt. The Max Planck Society decided against producing a ninth edition of the *Handbuch*. Instead, a *New Supplement* series was started in 1970. By the two-hundredth anniversary of his birth in 1988, there were 570 volumes (O'Sullivan, 1988), occupying fifty feet of shelf space, with about twenty volumes being added per year. Since the early 1980s the *Handbuch* has been produced in English. As with Beilstein's *Handbuch der organischen Chemie*, all data in Gmelin's *Handbuch* were critically evaluated.

Organization and Arrangement

As mentioned above, the first edition of the Gmelin handbook dealt with the forty-eight elements or "ponderable substances" known at the time. The meaningful ordering of these was one of Gmelin's major concerns (*Synergisms in Chemical Information*, 1992). First, he differentiated between inorganic and organic compounds. The ponderable inorganic compounds were then organized according to Gmelin's system, such that each volume deals with a different element. A classification scheme exists, with rules that determine in which volume compounds are located. Gmelin's system evolved into the present "principle of last position": Elements are assigned to one of seventy-one system numbers so that those that form anions are assigned lower numbers than those that form cations. The system numbers have no connection with atomic numbers. The information under each element is concerned with the element itself and all compounds that contain it along with other elements that have lower system numbers; for example, hydrogen chloride is discussed in volume 6 (chlorine), as H has System No. 1, and Cl, No. 6; $ZnCl_2$ is in volume 32 (zinc), and $ZnCrO_4$ is in volume 52 (chromium). Each section, or volume, can have supplements that update the work. Within each volume, the information is also arranged systematically: Analytical chemistry comes first, then atomic physics, ore preparation, chemical technology, electrochemistry, geochemistry, history, colloidal chemistry, coordination chemistry, corrosion and passivity, crystallography, economic deposits, metallurgy, mineralogy, physical properties, alloys, toxicity and hazards, and finally production statistics.

Chemisches Zentralblatt

At the age of twenty-nine Gustav Theodor Fechner (1801–87) conceived and edited *Chemisches Zentralblatt*, another early work. In 1817 Fechner matriculated at the University of Liepzig, where he remained for the rest of his life. He took his M.D. there, though he never practiced medicine. He also did research in physics and electricity and was appointed professor of physics in 1834. Later his research moved to psychology, for which he is primarily remembered (Gillespie, 1970–1980).

The *Pharmaceutisches Zentralblatt* began life in 1830. Its name changed in 1850 to *Chemisches und Pharmaceutisches Zentralblatt*. Six years later the name was shortened to *Chemisches Zentralblatt*. Indeed, the *Chemisches Zentralblatt* remained the most important abstracting service for chemistry globally until the World War II, which interfered with the production (Schulz & Georgy, 1994). Publication ceased for a period during the war, and the postwar publication is described by Dyson (1951) as "but a shadow of the pre-war publication." After the war, production was split between the West and East Berlin offices, which proved logistically difficult. The East Berlin office was closed in 1969, and in the same year publication finally ended. The West Berlin office along with Bayer AG continued to publish *Chemischer Informationsdienst*, an independent reference journal, which still continues today as the organic reactions database *ChemInform RX*.

Chemisches Zentralblatt was published weekly, with abstracts grouped under nine main headings, which were subdivided by use of a classification scheme. Initially, the plan was to cover German literature only. However, in 1919, coverage was expanded with the inclusion of the abstracts section of *Angewandte Chemie*, patents from all major industrial nations, and from 1926, notices. But coverage of German material was always superior to that originating in other countries, and abstracts were in German. To enhance its utility, the *Chemisches Zentralblatt* published cumulative indexes.

Beilstein's *Handbuch der organischen Chemie*

Like Gmelin's *Handbuch*, Beilstein's *Handbuch der organischen Chemie* is not an abstracting service but a secondary source that contains evaluated information. Beilstein's objective was to include only compounds and facts that were known to be reliable in terms of current scientific knowledge, so that the user is provided with a "concentrate" of the original literature free from errors and trivial or unvalidated information. Despite the huge growth in the literature, this was the aim of the publishers until recently. This made Beilstein's *Handbuch* different from traditional abstracts, which made no attempt to check the accuracy of the information included.

During the 1970s and 1980s the *Handbuch* became increasingly out of date in its coverage of the literature and suffered losses in its subscription numbers as a result of this, the huge subscription cost, and the slowness of its inevitable transition to publication in English. In the 1990s it is experiencing a resurgence in interest because of the *Beilstein CrossFire* service, now widely used in industry and academia.

Beilstein's Life

Friedrich Konrad Beilstein (1838–1906) was born in St. Petersburg to German parents (Gillespie, 1970–1980). At age fifteen Beilstein was sent to Germany, where he studied in Heidelberg under Bunsen and Kekulé. After two years he moved to Munich and studied under Liebig (Witt, 1909). He returned to Heidelberg where he became interested in organic chemistry. He then moved to Göttingen to study under Wöhler. There he worked on the cyanogen group, which led to his dissertation on murexide, for which he was awarded a doctorate in 1858 at the age of nineteen. In 1860 he was appointed Wöhler's assistant (privatdozent) and, by 1865, professor of organic chemistry. In 1866 he was chosen to succeed Mendeleev as professor at the Imperial Technological Institute of St. Petersburg (Witt, 1909), where he remained for the rest of his life. From 1865 to 1871 he edited *Zeitschrift für Chemie* (founded by Kekulé) along with Fittig and Hübner. In 1881 he was elected to the St. Petersburg Academy of Sciences, which gave him an independent income and laboratory. During the period from 1856 to 1889 Beilstein published more than one hundred experimental contributions to German and French journals (Huntress, 1938), and more in Russian. Over the years Beilstein sacrificed opportunities for original experimental work in order to continue his efforts to produce the *Handbuch*. The last seventeen years of his life were devoted entirely to its production.

Production of the Handbuch der organischen Chemie

Despite his significant research, Beilstein is most remembered for his *Handbuch*. The historical studies of Liebig and Wöhler and also the structural theories of Kekulé, van't Hoff, and Le Bel (stereochemistry) influenced Beilstein's work and provided the stimulus for the reinterpretation and reclassification of the known facts of organic chemistry (Luckenbach, 1981). These factors presumably inspired Beilstein to produce the *Handbuch* as well as his need to keep comprehensive records of the literature for his own research work. The first edition was published in 1881–83 and contained approximately fifteen thousand organic compounds, divided into five sections. The work comprised two volumes (2,200 pages). Only twenty-three journals were covered, and the reference list was twenty-three pages long. The first edition resembled a traditional textbook, with sections on organic analysis and determination of physical constants in addition to data on organic compounds. The publication was a success, selling out within a few months (Richter, 1938). The publishers (Leopold Voss) wished to produce a reprint, but Beilstein insisted on updating the work to create a second edition with corrections to errors and inaccuracies, including those resulting from an incomplete knowledge or understanding of the science when the first edition was published. The second edition consisted of three volumes (4,080 pages, 1885–1889), and the third edition, eight volumes (approximately 11,000 pages, including a supplementary series, 1892–1899; supplement, 1901–1906). The third edition was the last to be produced by Beilstein himself. Although Beilstein requested the help of other chemists in identifying inaccuracies, the production of the *Handbuch* was almost exclusively his work.

Prior to publication of the third edition Beilstein transferred responsibility for publication to the Deutsche Chemische Gesellschaft. Beilstein was concerned about maintenance of the quality of the *Handbuch* and in 1895 he authorized the publishers to approach Professor Paul Jacobson of Heidelberg to continue publication of the work, beginning with a supplement to the third edition. Jacobson, aware of the magnitude of the task, proposed that the *Handbuch* should be merged with the abstracts prepared for *Berichte* and the *Jahresbericht*, under the auspices of the Deutsche Chemische Gesellschaft. Although not entirely happy with this arrangement, Beilstein did agree to pass over his rights as author, and the directors of the society voted to continue the *Handbuch*. Jacobson became the editor to the supplement of the third edition, and Beilstein was satisfied that his *Handbuch* was in safe hands.

There was no formula index to the first three editions, but in 1884 M. M. Richter published his *Lexikon der Kohlenstoff-Verbindungen,* which served as an index to the third edition of the *Handbuch*. The index arranged substances in molecular formula order. Physical properties were also given, with references to the original literature and to the appropriate pages of Beilstein. The third edition of Richter, in four volumes, covered the literature up to 1909 and was superseded by the index to the fourth edition of Beilstein. In the Richter index formulae are divided into groups according to the number of carbon atoms present. Compounds are then subdivided based on the number of additional elements present. Formulae in each group are arranged in the order C,H,O,N,Cl,Br,I,F,S,P followed by others in normal alphabetical order. (See Mellon [1965] for a fuller account.) This order differs from the Hill system frequently adopted today, which lists C, then H, then all other elements in alphabetical order.

The fourth edition of the *Handbuch* commenced publication in 1918, with P. Jacobson and B. Prager as joint editors-in-chief. This edition covered the literature to the end of 1909. The task of scanning the primary literature had been partly removed because the staff was also working on *Chemisches Zentralblatt* (Richter, 1938). The staff checked the accuracy of the information for the abstracts against the original literature. This arrangement was successful for around twenty years, but an increasing number of editorial staff and their greater turnover, combined with the increase in the amount of primary literature, led to problems. With the second supplement of the fourth edition the Beilstein editorial staff reverted to consulting the original literature for about forty of the most important journals, referring to *Chemisches Zentralblatt* for the remainder. Compilation involved documenting the data in a strict order for each compound on a "slip": occurrence, formation, preparation, physical properties, chemical and biological behavior, analytical data, and salts. Each slip contained information from one paper only. The slips were assigned a system number according to the Beilstein classification scheme, which determined the position in the final handbook. In 1933 F. Richter took over responsibility from Beilstein. The fourth edition differs from the first three in scope and in the classification of the compounds. All compounds that had been synthesized, analyzed, and characterized were included and in addition any natural products that had been investigated. In total, thirty-one volumes were produced, twenty-seven covering the main classes of organic compounds, volumes 28 and 29 being the indexes (name and formula) and volumes 30 and 31 differing from the others in that they cover longer periods (vol. 30 to 1935 and vol. 31 to 1920) and that they cover natural products that had not been classified elsewhere or were poorly defined and therefore difficult to classify. Some classes of compounds were not dealt with adequately by Beilstein, for example, alkaloids. The fourth edition covered about 140,000 compounds. After this edition it was decided not to produce new editions but to bring out supplements. Table 1 gives the periods covered by the original work and supplements to Beilstein.

Table 1. The Supplements of the Beilstein Handbook

Series	Abbreviation	Period covered
Original work (Hauptwerk)	H	Up to 1910
Supplementary series I	EI	1910–19
Supplementary series II	EII	1920–29
Supplementary series III	EIII	1930–39
Supplementary series III/IV (Vols. 17–27 of supplements III and IV were combined as EIII/IV)	EIII/IV	1930–59
Supplementary series IV	EIV	1950–59
Supplementary series V	EV	1960–79

The work was produced in German up to the end of the fourth supplement; the fifth is in English. A cumulative set of name and formula indexes for the Hauptwerk and the first two supplements was published in 1955 and 1957.

Evolution of the Classification Scheme

There is a huge diversity of organic compounds, and this was also the case in Beilstein's time. Any classification scheme of organic compounds, therefore, needs to be able to accommodate all types of compounds and to be future-proof. The original scheme arose from existing knowledge of homologous series, parent nuclei, and functional groups, which have been known since about 1840, as well as differences between alicyclic and heterocyclic compounds discovered in the 1860s (Richter, 1938). It became clear to Beilstein that classification of the vast number of new compounds was becoming increasingly difficult according to the original scheme, mainly because of the increase in the number of heterocyclic compounds discovered in the 1880s. So the scheme was revised by Beilstein's staff before publication of the fourth edition.

The new scheme is a freely extendable method of classification, as new compounds can be incorporated on the basis of their structural features. There is a hierarchy that determines which structural feature has priority when compounds are allocated to system numbers and volumes. The scheme remains the same today, now dealing with over seven million compounds. The scheme is unique to the *Handbuch* and was the first such classification of organic compounds. As did Gmelin, Beilstein allocated system numbers to compounds. Specific volumes always covered the same range of system numbers. For example, 4-aminophenol has system number 1841 and is found in volume 13 and its supplements, which cover amines containing OH groups. The index to the Hauptwerk and the first two supplements gives the page numbers, and through the system number, its exact location can be identified in later supplements. The main divisions of the Beilstein classification are as follows: alicyclic compounds, volumes 1–4, and system numbers 1–449; isocyclic compounds, volumes 5–16, and system numbers 450–2358; heterocyclic compounds, volumes 17–27, and system numbers 2359–4720. As an example, volume 24 covers heterocyclic compounds containing two nitrogen atoms, which also contain an oxo group; their system numbers fall in the range 3555–3633. Many chemical literature guides give further details of the present classification scheme: See, for example, Skolnik (1982) and booklets produced by the Beilstein Institute.

British Chemical Abstracts and *Analytical Abstracts*

British Chemical Abstracts had its origins in 1849, when the Chemical Society began to publish abstracts (Whiffen, 1991). The Society of Chemical Industry followed with its abstracts in 1882. Eventually it was decided that the overlap in coverage justified a merger, which occurred in 1926, leading to publication of *British Chemical Abstracts*. This merger resulted in formation of a bureau that produced *Abstracts A (Pure Chemistry)* and *Abstracts B (Applied Chemistry)*. In 1937, the *A* abstracts were split into three sections: *Ai Pure chemistry (General, Physical and Inorganic)*, *Aii Pure Chemistry (Organic),* and *Aiii Pure Chemistry (Biochemistry)*. In 1938 the publication was renamed *British Chemical and Physiological Abstracts* after the Physiological Society joined the bureau, which was renamed the Bureau for Chemical and Physiological Abstracts. Later the Society for Experimental Biology joined in production of the publication. The *B* section was divided as follows: *Bi General and Inorganic Chemistry*, *Bii Industrial Organic Chemistry*, *Biii Agriculture, Foods, Sanitation* (from 1938). The Society for Analytical Chemistry had published abstracts in its primary journal, *The Analyst*, but it was decided to publish these as part of *British Chemical Abstracts* and a new section *C Analytical Chemistry* was introduced in 1944.

World War II caused problems with production because of the lack of published scientific work, a paper shortage, and the lack of access to many European journals. In 1945 the bureau was renamed the Bureau of Abstracts and the journal was renamed *British Abstracts*. The abstracts were never financially viable and were discontinued in 1953, with debts to the tune of around £90,000, which were paid by the chemical industry, the societies involved with its production, and the British government. The *Journal of Applied Chemistry* assumed responsibility for abstracting the applied literature.

One positive outcome from these events was the birth of *Analytical Abstracts*, which began in 1954. Having decided to stop publishing abstracts in *The Analyst*, the Society for Analytical Chemistry still wanted to communicate abstracts to analytical chemists. *Analytical Abstracts* is still published today and is profitable. Differences in material covered (for example, standards) and the ability to search for compounds in specific roles in the electronic version, such as analyte or matrix, complement the coverage of *Chemical Abstracts* for the analytical chemist.

Chemical Abstracts

The major abstracting service in chemistry that dominates today is *Chemical Abstracts* (*CA*), which began in 1907. By then there were more than sixty abstract journals in pure science (Manzer, 1977; Skolnik, 1982). *CA* was born partly out of American chemists' dissatisfaction with the coverage of American chemical literature by European abstracting journals (Baker, Horiszny, & Metanomski, 1980). This dissatisfaction came despite the trend for abstracting journals to broaden their coverage to include literature from countries other than their country of origin. Faculty members at MIT tried to remedy this by producing *Review of American Chemical Research* in 1895, the forerunner to *CA*. *CA* was sponsored by the American Chemical Society, its first editor being W. A. Noyes, Sr. The first issue of *CA* contained 11,847 abstracts (Wolman, 1988), taken from 396 journals (Donnell, 1995). The number of journals covered increased to a thousand by 1922 and two thousand in 1932; today the number is around nine thousand, along with patents from twenty-seven patent offices.

Subject Coverage

CA's mission was to abstract the complete world's literature of chemistry, at first glance a straightforward objective, but in fact one that led to problems with the definitions of three words, *complete*, *abstract*, and *chemistry*. E. J. Crane (1889–1966), editor of *CA* from 1915 to 1958, indicated that publications suitable for inclusion were "studies of new chemical reactions, new information on known reactions, chemical, physical and biological properties of elements or compounds, apparatus of particular interest to the chemist or chemical engineer and procedures that in themselves may not involve chemistry but are essential to an industry that is generally considered chemical." *CA* covered applied and industrial chemistry as well as pure chemistry from the outset. This was encouraged by an early worker on the publication, W. Russell Stemen. He was also aware of the importance of patents as an information source, although patent summaries were brief until 1945.

In the first volume there were twenty-four issues of *CA*. Each issue was divided into thirty sections. The biggest sections were organic chemistry and biological chemistry. The patents section covered U.S., British, French, and German patents at the outset. Early decisions about the classification scheme laid the foundations of the systems in use today. Present users of printed *CA* know that the section headings have necessarily evolved and expanded to reflect changes in the importance of research areas and the appearance of new subjects. However, it was not until 1962 that a major overhaul of the classification took place, when the number of sections was increased to seventy-three. By 1980 there were eighty sections.

The Abstracts

The first issue of volume 2 contained an informative section titled "Organization, Directions for Assistant Editors and Abstractors and List of Journals," which provides insight into the selection procedures, editorial policy, and coverage of *CA* at the time. The duties of an assistant editor are indicated to have been:

> (1) The selection of the abstractors for his division. (2) To select the journals which contain material important for the division and to see that no such journals are overlooked. (3) To keep an oversight of the character of the abstracts and make sure that they give an adequate report of the articles abstracted, in good English and with the necessary brevity. (4) To advise the editor with regard to defects in *CA* and to indicate directions in which the journal may be improved. (5) To make sure that abstracts are prepared for all journals and articles assigned to his care. (6) To examine the proof for his division.

Each section had its own assistant editor. On editorial policy, in 1917, the "guidelines for abstractors" instructed abstractors not to make any personal judgments of the content of the papers being abstracted; this was the responsibility of the reader of the information. This is in contrast to the policy used in the production of Beilstein and Gmelin.

In 1907 about 50 percent of the abstracts were of articles originally published in German. By 1937 this had been reduced to 15 percent, with 40 percent in English, 5 percent in Japanese, 7 percent in Russian, and 27 percent in other languages. Now about 80 percent are in English, with only 2 percent in German, demonstrating the shift from German to English as the principal language for publication about chemistry. By 1959 papers were being received in fifty different languages (Heumann & Bernays, 1959), which gave rise to problems with translation of the subject matter and with transliteration of authors' names from Chinese, Cyrillic, and other alphabets.

Indexes

The first volume of *CA* had author and subject indexes, which occupied 363 pages. The number of pages taken up by the indexes increased with time as would be expected. The subject index covered both chemical names and general subjects. In 1907 the original subject index to volume 1 contained 7,850 index headings (Zaye, Metanomski, & Beach, 1985), and about 19,000 subject index entries, equivalent to 0.6 percent of the number of entries in 1983. Of the 7,850 headings 75 percent had only one reference associated with the index term, the maximum number being around 140. From the outset the overall aim was to index abstracts by subject rather than word; index headings were controlled, with cross references guiding the user to the correct heading. The headings have been revised as the need has arisen; some headings are the same today as they were in 1907. The development of indexing policy was initiated by Austin M. Patterson, the *CA* editor from 1909 to 1913, but as in many other areas, the key player was E. J. Crane, who laid the foundations of today's indexing system. Chemical nomenclature experts ensured consistency among the chemical names used, and naming conventions evolved in the same way as general subject headings. The decennial indexes required complete re-editing of the annual indexes. Bernier and Crane (1948) state that "chemistry is a growing science, and the indexer of an abstract journal must frequently deal with nomenclature in its early stages when lack of standardization and even lack of full knowledge make for indefiniteness. It is on this account that the collective indexing presents so many tough problems." In 1972 the subject index was divided into the general subject index and the chemical substance index, with general topics indexed in the former, specific compounds in the latter. The chemical substance indexing scheme involves inversion of names so that related compounds are grouped under the same heading.

Formula indexes were first published in 1920. They are organized according to the Hill system (Hill, 1900, 1907), which lists elements in a compound in the order carbon, then hydrogen, followed by all other elements in alphabetical order. For example, 2-nitropyridine would have the formula $C_5H_5N_2O_2$. These indexes were not intended to be used independently of the subject indexes; they provide names of commonly referenced compounds, and for these "common" compounds, the references themselves are to be found in the subject index. The subject index includes modifying phrases (not included in the formula indexes) to aid in determining the usefulness of references. The formula index leads directly to abstract numbers for "uncommon" compounds, which would usually have only one or a few abstracts associated with them in a decennial index period. There was a somewhat arbitrary cutoff of greater than fifty references differentiating between "common" and "uncommon" compounds.

Patent indexes started in 1912. A further index is the Index of Ring Systems, which first appeared as a separate publication with the 1957–66 Collective Index (it was formerly part of the introduction to the subject index).

The Production Process

In 1907 there were 129 volunteers involved in the production of *CA*, who received no remuneration (Baker, Horiszny, & Metanomski, 1980). In 1929 minimal pay was offered. By 1938 there were over four hundred abstractors in the United States and elsewhere (Scott, 1938); by 1954, over a thousand; and by 1961, more than three thousand. The Columbus-based staff coordinated this activity, final responsibility resting with E. J. Crane. The Columbus editors were also responsible for assigning each abstract to the appropriate section and cross-referencing. Scott (1938) states that while she was employed at *CA*, there were twenty-five full-time workers, and their responsibilities also included editing manuscripts from the abstractors, other aspects of quality control, and indexing.

After 1966 the policy of employing volunteers was gradually phased out, so that only 8.6 percent of abstracts were compiled externally, the number of abstractors falling from 3,292 in 1966 to about 1,000 in 1979. This decrease was mainly caused by difficulties of administering such a large number of dispersed personnel and also the arrival of computers in the 1960s, enabling automation of some production procedures that could then be carried out in house.

An account of the compilation of the indexes is given by Bernier and Crane (1948). The indexers, all chemists, worked from pages of typeset abstracts. Words to be included were underlined or noted in the margin. An index card was made, with the index modification and the volume and column reference. Cards were checked to eliminate inconsistencies. Over the years the depth of indexing increased; For example, the number of index entries increased from 2.9 per abstract in 1936 to 4.2 in 1946.

Scott (1938) perceives that certain personal qualities are needed for the type of work at *CA*:

> Accuracy should perhaps rank the highest and with that conscientiousness, patience, a meticulous attentiveness to detail, . . . power of concentration, good judgement, an interest in words as words, a love of puzzles . . . The analytical rather than the creative type is probably best suited for the work . . . Work of this type is not for the overly energetic or restless person.

Patents

Use of patents as an information source has always been overshadowed by other forms of publication (Schofield, 1996), but patents cannot be ignored when a literature search is conducted; the figure of about 80 percent of information in patents never appearing elsewhere is frequently quoted. Publications that specialize in producing summaries of patents have existed for the last two centuries. One example is the *London Journal of Arts and Sciences and Repertory of Patent Inventions*, which was compiled by W. Newton and concentrated on civil and mechanical engineering inventions. Some examples of publications in chemistry and chemical engineering are Kirk and Othmer (1947–56); *Fortschritte der Teerfarbenfabrikation und verwandter Industriezweige*, covering 1877 to 1942 and, although concentrating on German patents, also covers others in the later years and deals with dyes and related subjects; *Fortschritte der Heilstoffchemie, erste Abteilung: Das deutsche Patentschriftwesen*, published from 1926 to 1939, covering German patents on medicinals, cosmetics, and other aspects of organic chemistry; *Fortschritte in der anorganisch-chemischen Industrie*, describing German patents from 1877 to 1932; *Zusammenstellung der Patente auf dem Gebiete* *der organischen Chemie*, which discusses organic chemistry patents from 1877 to 1905; *Chemical Patents Index,* which includes all U.S. chemical patents granted from 1915 to 1924. National patent offices also publish abstracts and alerting services, such as the U.K.'s *Official Gazette (Patents)* and the U.S.'s *Official Gazette of the United States Patent and Trademark Office*. All organize their patents according to classification schemes and are indexed. The importance of patents to the chemical industry was recognized by most of the producers of secondary sources described, and their coverage has been mentioned. There was considerable variation in the comprehensiveness of subject and country coverage. Derwent Information is now a major producer of patent information retrieval tools, first with their *Patents Abstract Publications,* which began in 1951, and now with the *World Patents Index* database.

Conclusion

The need for secondary sources in chemistry arose in a way similar to that in other disciplines, that is, when the primary literature of the subject started to become unmanageable. For chemistry this was perceived earlier than for most other subjects. The most important type of secondary source is the abstract journal, devoted exclusively to publication of abstracts, themselves having origins considerably before such journals.

A major difference between chemistry and other disciplines is the importance of chemical structures as the universal means of communication, necessitating multiple access routes to information. This is usually achieved through formulae or chemical names, often in combination with a substance classification scheme, which has led to the evolution of especially sophisticated information retrieval systems in chemistry.

Why did secondary sources develop to a greater extent in some counties than others? Many secondary sources were conceived during the nineteenth century, when the primary literature was growing because of the flourishing chemical industry and research activity in universities and research organizations. The Gmelin and Beilstein handbooks commenced during a particularly productive period for research in Germany. In the aftermath of the Napoleonic wars, the Prussians were keen to develop a counter to French culture. One result of this nationalistic response was a reform of science in German universities, giving a great stimulus to research. In Great Britain, science was less organized, and science education somewhat weaker at that time. *CA* was developed during a period of growth in the chemical industry in the United States.

So why did some secondary sources survive while others did not? To some extent it was survival of the fittest, but in addition wars and cultural developments had an effect. The flexibility of the classification schemes employed in the handbooks of Gmelin and Beilstein have contributed to their survival. These schemes revolve around functional groups, ring systems, and the chemi-

Important Secondary Sources in Chemistry

Some significant works not specifically mentioned in the text are also included.

❖ = original title
➢ = publication related to that above, name changed

- ❖ *Crell's Chemische Journal für die Freunde der Naturlehre,* 1778–1781
- ❖ *Chemisches Annalen für die Naturlehre, Arzneygelahrtheit, Haushaltungskunst und Manufacturen,* 1784–1803
- ❖ *Neues Chemisches Archiv,* 1784–1791
- ❖ *Beiträge zu den Chemischen Annalen,* 1785–1799
- ❖ *Gmelin's Handbuch der Anorganische Chemie,* 1817–date
- ❖ *Pharmaceutisches Zentralblatt,* 1830–1849
  - ➢ *Chemisches und Pharmaceutisches Zentralblatt,* 1850–1856
  - ➢ *Chemisches Zentralblatt,* 1856–1896
  - ➢ *Chemisches Zentralblatt,* 1897–1969
  - ➢ *Chemischer Informationsdients* (*ChemInform*), 1970–date
- ❖ *Répertoire de Chimie Pure,* 1858–1863
- ❖ *Répertoire de Chimie Appliquée,* 1858–1864
- ❖ *Beilstein's Handbuch der Organischen Chemie,* 1881–date
- ❖ *Review of American Chemical Research,* 1895–1906
  - ➢ *Chemical Abstracts,* 1907–date
- ❖ *Houben-Weyl: Methoden der organischen Chemie,* 1909–date
- ❖ *British Chemical Abstracts,* 1926–1937
  - ➢ *British Chemical and Physiological Abstracts,* 1938–1944
  - ➢ *British Abstracts,* 1945–1953
  - ➢ *Analytical Abstracts,* 1954–date
- ❖ *Nippon Kagaku Soran,* 1927–1957
- ❖ *Theilheimer's Synthetic Methods for Organic Chemistry,* 1946–date
- ❖ *Bulletin Signalétique,* 1940–date
- ❖ *Referativnyi Zhurnal Khimiya,* 1952–date
- ❖ *Index Chemicus,* 1960–date
  - ➢ *Current Abstracts of Chemistry,* 1970–date

cal elements, that is, the true language of chemistry. The majority of these structural features were known before the publications originated, and so the classification schemes have needed little amendment as new structures can be slotted into the existing schemes. In addition the financial support of these publications by the German government clearly helped.

A common factor among some of the publications that failed to survive was the inadequacy of the coverage of the international literature. After World War II and during the rise of *CA*, scientific research in the United States boomed, whereas Europe was still recovering from the war. It was dangerous for the producers of secondary sources to give scant coverage to U.S. publications. The decline in *Chemisches Zentralblatt* and *British Chemical Abstracts* can be attributed in part to this factor. *CA*'s breadth of coverage, both in terms of subject and country of origin of publications, has reduced the impact of most other surviving abstract services, even in their country of origin. The importance of international coverage is confirmed by the success of *Science Abstracts*, a British publication that has survived where *British Chemical Abstracts* failed, since the former paid more attention than the latter to the American literature from its inception.

Another factor in survival is specialization. For example, *Analytical Abstracts* survives because it covers sources of particular interest to analytical chemists, such as standards, not included in *CA*.

It seems clear, however, that individuals like Beilstein, Gmelin, and Crane, who laid the foundations of today's information retrieval systems in chemistry, will continue to have an influence on the organization and evolution of secondary information sources in chemistry while chemistry as a subject continues to develop and evolve.

References

Baker, D. B., Horiszny, J. W., & Metanomski, W. V. (1980). History of abstracting at Chemical Abstracts Service. *Journal of Chemical Information and Computer Science, 20*(4), 193–201.

Bernier, C. L., & Crane, E. J. (1948). Indexing abstracts. *Industrial and Engineering Chemistry, 40*(4), 725–730.

Bowman, C. M. (1974). The development of chemical information systems. In J. E. Ash & E. Hyde (Eds.), *Chemical information systems (*pp. 6–12). Chichester, England: John Wiley.

Collison, R. L. (1971a). *Abstracts and abstracting services.* Santa Barbara/Oxford: ABC-Clio Press.

Collison, R. L. (1971b). *The annals of abstracting 1665–1970.* School of Library Science and the University Library, University of California, Los Angeles.

Cooper, M. (1982). Secondary information services in science and technology: A wide-angle view. *Journal of the American Society for Information Science*, *33*(3), 152–156.

Crosland, M.P. (1962). *Historical studies in the language of chemistry.* London: Heinemann.

Crosland, M. P. (1994). *In the shadow of Lavoisier: The* Annales de Chimie *and the establishment of a new science*. British Society for the History of Science Monograph No. 9. Oxford: BSHS.

Donnell, E. P. (1995). Growth and change in the world's chemical literature as reflected in *Chemical Abstracts. Publishing Review Quarterly*, *10*(4), 38–46.

Dyson, G. M. (1951). *A short guide to the chemical literature*. London/New York/Toronto: Longmans, Green and Co.

Gillespie, C. C. (Ed.) (1970–1980). *Dictionary of scientific biography*. New York: Charles Scribner.

Heumann, K. F., & Bernays, P.M. (1959). Fifty foreign languages at *Chemical Abstracts. Journal of Chemical Education, 36*(4), 478–482.

Hill, E. A. (1900). On a system of indexing chemical literature; adopted by the classification division of the U.S. patent office. *Journal of the American Chemical Society*, *22*(8), 478–494.

Hill, E. A. (1907), The chemical card index of the patent office. *Journal of the American Chemical* Society, *29*(6), 936–941.

Huntress, E. H. (1938). 1938: The one-hundredth anniversary of the birth of Friedrich Konrad Beilstein (1838–1906). *Journal of Chemical Education, 15*(7), 306.

Kaser, R. T. (1995). Secondary information sources: Mirrors of scholarly communication. Forces and trends. *Publishing Research Quarterly*, *11*(3), 10–24.

Kirk, R. E., & Othmer, D. F. (1947–1956). *Encyclopedia of chemical technology*. New York: Interscience.

Luckenbach, R. (1981). The Beilstein Handbook of Organic Chemistry: The first hundred years. *Journal of Chemical Information and Computer Sciences*, *21*(2), 82.

Manzer, B. M. (1977). *The abstract journal, 1790–1920*. Metuchen, NJ: Scarecrow Press.

Mellon, M. G. (1965). *Chemical publications: Their nature and use.* 4th ed. New York: McGraw-Hill.

Neufeld, M. L. & Cornog, M. (1983). Secondary information systems and services. *Annual Review of Information Science and Technology*, *18*, 151–183.

Oesper, R. E. (1949). Erich Pietsch. *Journal of Chemical Education, 26*(5), 251–252.

O'Sullivan, D. A. (1988, August 15). Gmelin handbook flourishes at 200th birthday of founder. *Chemical and Engineering News,* 22–24.

Richter, F. (1938). How Beilstein is made. *Journal of Chemical Education*, *15*(7), 310–316.

Schofield, H. (1996). Patent information usage by chemists in universities. In H. Collier (Ed.), *Proceedings of the 1996 International Chemical Information Conference*, Nîmes, France, (pp. 111–120). Calne, UK: Infonortics.

Schulz, H. & Georgy, U. (1994). *From CA to CAS online*, 2nd ed. Berlin: Springer-Verlag.

Scott, J. D. (1938). My work with *Chemical Abstracts. Journal of Chemical Education, 15*(7), 271–275.

Skolnik, H. (1977). Historical aspects of patent systems. *Journal of Chemical Information and Computer Sciences*, *17*(3), 119–121.

Skolnik, H. (1982). *The literature matrix of chemistry*. New York: Wiley.

Synergisms in chemical information. (1992). Berlin: Springer-Verlag.

Tate, F. A. (1967). Handling chemical compounds in information systems. *Annual Review of Information Science and Technology, 2*, 285–309.

Walden, P. (1954). The Gmelin chemical dynasty. *Journal of Chemical Education*, *31*(10), 534–541.

Whiffen, D. H. (1991). *The Royal Society of Chemistry: The first 150 years*. London: RSC.

Whittingham, D. J., Wetsel, F. R., & Morgan, H. L. (1966). The computer based subject index support system at Chemical Abstracts Service. *Journal of Chemical Documentation, 6*(4), 230–234.

Williams, T. (1977). Scientific literature: Its influence on discovery and progress. *Interdisciplinary Science Reviews, 2*(2), 165–172.

Witt, O. N. (1909). Obituary notices: Friedrich Konrad Beilstein. *Journal of the Chemical Society*, 1646–1649.

Wolman, Y. (1988). *Chemical information: A practical guide to utilization*, 2nd ed. Chichester, England: Wiley-Interscience.

Zaye, D. F., Metanomski, W. V., & Beach, A. J. (1985). A history of general subject indexing at Chemical Abstracts Service. *Journal of Chemical Information and Computer Science, 25*(4), 392–399.

The History of Managing Technical Information at DuPont

Florence H. Kvalnes

Abstract

In the 1960s, DuPont developed computer databases to manage its collection of technical information, and many of the individual departments had their own information centers.

During this time the duplication of efforts between departments, especially in the area of patent services, was studied. These studies had an impact on the handling of proprietary technical information. DuPont then developed, built, and implemented an integrated system for the storage, retrieval, and distribution of its in-house scientific and technical information. In addition to changes in the mechanical handling of information, changes were made in the way the content was described, for example, from classification systems to concept coordination. Many of the fundamental principles for storage and retrieval of technical information developed in the 1960s, including representations of atom-bond-atom structure and a thesaurus, were adopted and are still in use. These innovations resulted in the Scientific Corporate Information Online (SCION) database created in the early 1990s, which provides online access for DuPont's scientific community to proprietary technical information. The next generation of the SCION database is under development and will take advantage of new computer and information science technologies. An intranet was developed and continues to grow and gain importance in managing DuPont's information.

Introduction

Technical information services, although they are mostly decentralized, have been a tradition at DuPont for nearly a century. The first formal libraries were established in 1917–18 (Duncan, 1951), and groups staffed with scientists and engineers were gradually formed within libraries to index patent and proprietary information in the form of formal research reports. These groups evolved into information centers that provide a variety of information services, depending on the needs of the organization.

This article focuses on the tools and techniques for describing proprietary technical information and the development of computer databases for managing this information. For purposes of this discussion, proprietary technical information is defined as that which is documented in formal reports. It is not a comprehensive history of managing all technical information at DuPont.

Corporate Culture

DuPont company policy and corporate culture have contributed to the decentralization of information services. The company is divided essentially into autonomous segments or business units and a few staff functions, such as corporate information science, legal, financial, and so forth. The business units are not obligated to employ the services of a centralized information service organization. Thus, there has been no unified approach to handling information at DuPont (Conrad, 1955). Information technology infrastructure has a similar organizational history. Many business units or departments had their own computers and computer groups. For a long time the only centralized computer group served corporate financial, accounting, and human resource requirements.

The business units are built around a very diverse group of products, such as fibers, specialty chemicals, and agrochemicals, among others. These diverse scientific and business interests have contributed to the various challenges for managing information (Figure 1).

New directions in technology mean that information scientists need to keep up to date with current technologies as well as learn new ones as the company diversifies into new areas. Information scientists must also learn to use a variety of tools for managing information.

Figure 1.
Diversity of Information

- Scientific—chemistry, physics, math, life sciences
- Engineering
- Manufacturing
- Technical marketing services
- Market research
- Environmental science

Transition to a global company has meant networking information services overseas and adjusting to different languages and attitudes about how to manage information.

The 1960s marked a departure from the policy of little direct intervention in the management and use of information services by corporate level management. The company's executive committee requested a study of the duplication of effort in ordering and indexing patents and in indexing and searching technical information recorded in formal research reports. In 1964 this study resulted in the consolidation of nine separate information groups to form two groups called the Central Report Index and Central Patent Index (Rasmussen & Van Oot, 1969). Further consolidation occurred in 1985, when there was some centralization of library administration and services with the creation of the Technical Library Network. Also in 1985, the Technical Library Network was combined with the Central Report Index, Central Patent Index, and Language Services (Nichols, Sikes, Isselman, & Ayers, 1995–1996) to create the Corporate Information Science organization.

As a rule, the staffs of DuPont's information centers have been and continue to be educated in science or engineering. Many information professionals also have degrees in library or information science. Their responsibilities include conceptual analysis, searching, library services, patent searching, and investigating new ways to manage information or knowledge. These are recognized as the core competencies of the organization (Ayers, 1994–1995; Nichols et al., 1995–1996).

Conceptual Analysis

Between 1917–18, when the first DuPont libraries were formed, and about 1958, technical information in libraries, file rooms, and information centers was "indexed" using various classification schemes. In the mid-1950s it was becoming apparent that because of the increase in information to be classified, new ways of managing this information were needed. At that time the charter of a group of consultants in the Business Analysis group of the engineering department was to find better methods for managing engineering information. They extensively studied and evaluated the methods of E. Wall (1959), M. Taube (1953, 1962), F. W. Lancaster and J. Mills (1964), H. P. Luhn (1957), G. Salton (1961), and S. Herner, F. W. Lancaster, and W. F. Johannigsmeirer (1964). Most if not all of the techniques summarized in a recent publication by F. G. Kilgour (1997) were tried or used at one time or another (Dinwiddie & Conrad, 1954; Edge, Fisher, & Bannister, 1957) (Figure 2).

Information centers in other parts of the company learned of the efforts of the Business Analysis group of the engineering department and asked for assistance to improve the centers' methods of handling information. Concept coordination, using uniterms and a thesaurus, was first adopted about 1958 for indexing technical reports, engineering drawings, patents, and correspondence (Costello, 1961). Dual dictionaries were printed and distributed to many of the DuPont plants, laboratories, and construction sites for on-site searching by chemists and engineers.

The Business Analysis consultants looked for ways to solve the problems created by different points of view about the significance of information in a document and how these points of view would affect retrieval. They developed a word association matrix to help in identifying indexing terms, and they adopted links and roles to indicate syntax. Links were used to group sets of indexing terms together to reduce false correlation, and roles were used to designate word order and relationships between terms within a link.

Figure 2.
Conceptual Analysis

- Pre-1950s: Various classification schemes
- 1950s–1960s: Active study, investigation of developments in information storage and retrieval
- About 1958: Uniterms and pre-coordinated vocabularies adopted
- Dual dictionaries, McBee keysort cards, optical incidence cards, mechanized card sorting

Word Association Matrix

The word association matrix was a listing of lead terms and the frequency of association in documents with the terms listed under the lead term. This was an attempt to

Table 1. Word Association Matrix

Lead Term	Frequency	Association
AIR POLLUTION	50	
Contaminating (see also impurities)	50	100%
Air (see also atmospheres)	48	96%
Ashes	13	26%
Power plants—power houses	8	16%

help not only indexers but also searchers find terms to search without regard to point of view.

Using Table 1 as an example, *air pollution* is the lead term, and its frequency of use in documents is fifty. *Contamination* is used in association with *air pollution* in indexing the same fifty documents; therefore, the association is 100 percent. *Air* was used with *air pollution* to index forty-eight of the same documents; thus, the association is 96 percent, and so on.

Links. Links are used to accumulate indexing terms into a sentence-like association, which describes information about a concept. When properly applied, links result in reduction of false retrieval. However, word order is not established, since common practice is to alphabetize the terms within the link.

For example, a document discusses two subjects: the steam cleaning of autoclaves and the design of extruders. The indexing terms are divided into link A and link B, as shown in Figure 3. This reduces unwanted retrieval for searches for the design of autoclaves or for steam cleaning of extruders because the terms are separated into different units. Links are taken into consideration when Boolean logic is performed.

Figure 3.
Use of Links

Link A	*Link B*
Autoclaves	Design
Cleaning	Extruders
Steam	

Roles. Another device developed by the Business Analysis consultants in an attempt to give syntax to the selected terms within a link was roles. There were twelve roles in the original set, and they were assigned to every indexing term. An internal study showed that most were ineffective except for those that described chemical reactions (Van Oot, Schultz, McFarlane, Kvalnes, & Riester, 1966) (Figure 4).

Figure 4.
Role Definitions

Role	*Definition*
1	Using, by means of, by
2/9	Cause and effect
6	Reaction by-product, impurity
8	Major topic
10	Design, drawing of
11	Receiving a physical modification
12	Claim or disclosure of in a patent

When information groups from nine of the departments were consolidated into the Central Report Index in 1964, concept coordination continued as the method of indexing. Links and roles were implemented for all documents. The various role sets were consolidated. The above roles were dropped, that is, converted to 0 for "Other," since indexing practices required that all terms be indexed with a role. Only the ones used for indexing chemical reactions were retained (Figure 5).

Figure 5.
Role Definitions—For Chemical Reactions

Role	*Definition*
3	Reactants in a chemical reaction
4	Special agent in reaction, e.g., catalyst
5	Reaction medium, atmosphere
7	Products of a chemical reaction
0	Other, includes properties of, uses of, etc.

Development of a Thesaurus

The various individual information centers in the old decentralized system all considered a thesaurus essential for their indexing and retrieval operations; some thesauri were synonym lists and others were hierarchical. The earliest hierarchical thesaurus at DuPont seems to have been developed by the engineering department about 1960. It was derived from their word association matrix, which was based on the statistical occurrence of terms used in the indexing of individual reports.

With the consolidation of nine different departments into the Central Report Index in 1964, a common vocabulary had to be developed. Two indexers (one a chemist, the other an engineer) experienced in term editing and hierarchical thesaurus preparation were assigned the task of consolidating the nine different vocabularies into one. About 28,000 different general terms were reviewed for selection of word form, elimination of synonyms, and significance of relationships and cross-terms. The final product consisted of 11,000 uniterms, of which 5,000 were names of equipment, techniques, devices, processes, or properties. The other 6,000 were trade names. This editing effort required eighteen person-months. The previous nine departmental indexes, that is, the indexing terms with their respective documents, were converted to the new vocabulary to produce a single index.

Between 1964 and about 1972, it became apparent that the uniterm form of a controlled vocabulary was not very effective for indexing documents covering the diverse areas of technology that were of interest to the company. A review of problem areas, notably in the area of properties, was undertaken. A decision was made to use pre-coordinated terms, such as *light stability* instead of using *light* and *stability*, and to add qualifiers to divide some terms, such as *growth* into *growth, biological* versus *growth, of markets.*

Another area of difficulty in searching resulted from practices used in indexing DuPont product lines, especially products that were all made by the same processes but that differed in certain parameters, such as use or size. For example, Dacron polyester fiber is spun from a specific polymer and is used in many kinds of apparel and bedding. The set of indexing terms used to index this product line describes a large number of polyester products. The sizable retrieval from searches for information on one of these products resulted in screening many abstracts to select those relevant to a query. From such experiences with too many retrievals or false correlations, it was decided to review the thesaurus and to create pre-coordinated terms in the technology areas where the efficiency in indexing and searching would be improved.

Chemical Structure

The studies by the Business Analysis group that led to the adoption of concept coordination showed the need for new ways to store and retrieve chemical structures and information. The nine information groups did not index chemical substances the same way. The methods included Chemical Abstracts Service nomenclature rules, International Union of Pure and Applied Chemistry (IUPAC) nomenclature rules, a system based on Beilstein classification, and fragmentation systems. These systems were error prone and costly and did not permit easy retrieval.

In 1962, two DuPont engineers, Donald Gluck and Leslie Rasmussen, developed GRAM, the Gluck Rasmussen association matrix system. GRAM was capable of handling all chemicals, including polymers, of interest to DuPont (Gluck, 1965). Two unique algorithms permitted reasonable computer costs. The input algorithm was based on uniquely positioning each atom in an ordered list according to the atoms to which it was connected. A second algorithm was used to transform the ordered compounds into a nonredundant compact list for storage and search. Another major advantage over the other systems known at the time was that input from drawn structures could be accomplished without any rules for numbering the atoms. Therefore, someone not trained in chemistry could input structures.

Between 1962 and 1964 there were discussions with Chemical Abstracts Service (CAS) about the development and use of topology for storing and searching chemical structures by atom-bond-atom via connection tables ("New system," 1963). GRAM was shared with the American Chemical Society for use by CAS in developing a chemical registration system for the entire chemical industry. Then CAS and DuPont agreed to work together on the development of a chemical registration system ("CAS and DuPont," 1964). This collaboration involved ordering of tables of atoms and bonds in the registry process, defining the compound types to be handled and the methods for handling exceptions, defining a method of file organization and screen generation, and coming up with search logic techniques and economics. H. L. Morgan, of Chemical Abstracts Service, generated a unique machine description for chemical structures using the algorithms developed by Rasmussen and Gluck (Morgan, 1965). In 1964 CAS shared the input and the atom-by-atom search programs with DuPont for testing.

For its own system, DuPont decided to develop computer-generated screens for searching based on the fragments used in several of the earlier search systems and to continue its own methods for registering polymers. The DuPont method makes certain assumptions about polymer structure to collect similar references at one point (Schultz, 1975). CAS indexes the polymer description available in the source document and does not make assumptions based on the author's description. This practice can result in the indexing of information about the same polymer under more than one reference point or registration number. Several references describe the differences between the CAS method of registering polymers and the DuPont method in more depth (Patterson, Schultz, & Wilks, 1995; Schultz & Wilks, 1997; Wilks, 1997a–d).

Computer Systems

Batch Systems

During the 1960s manual systems were converted to several small computer systems. These were searched and updated at scheduled intervals (batching), not on demand as is possible with current technology. In 1961 the management of some of the information centers explored the joint development of a computer program for information storage and retrieval. The program was written so that each department could use the same storage and retrieval functions but maintain separate databases. The Multidepartment Information Retrieval System, as this system was named, was programmed for an IBM 705 and completed in 1962. The information in the file included indexing terms and report numbers in an inverted file format. There was also a hierarchical thesaurus for indexing. Documents were posted to all upper-level terms or more generic terms in the hierarchy. This enabled searching for a family of documents without having to “or” many terms together to collect all of the documents in a given hierarchy. About 1962 nine departments were using this computer information storage and retrieval system and concept coordination. Consolidation of these nine indexes into one was facilitated by the use of the Multidepartment Information System in 1964 when the Central Report Index was formed.

With the decision to operate a central index for technical report storage and retrieval (Montague & Schirmer, 1968), work began on a system that would upgrade the computer system from the IBM 705 to the newer IBM 1410/7010 computer (Hoffman, 1968). The system on the IBM 1410/7010 computer was an interim system. The goals for developing this system were 1) to consolidate the separate databases into one database rather than maintain nine separate databases; and 2) to allow time to develop the requirements for a more comprehensive system based on the needs of the searchers. Online searching had not yet been introduced, and searching for proprietary technical information was done by the staff of the Central Report Index for company scientists and engineers. Increased efficiency and reduced costs for the newly created Central Report Index were achieved by being able to search one database rather than nine separate databases.

The database retained the inverted file structure used in the IBM 705 system. There were two inverted term-document files, one for compound registration numbers and the documents posted to them and the other for thesaurus terms and their documents. Integrated with these files was the Chemical Structure Storage and Search System (CS4) registry file (Hoffman, 1968) and a hierarchical thesaurus. CS4 stored the topology of chemical structures in the form of connection tables and served as a second-level index to the primary document system files. Registration numbers were assigned based on determinations of the uniqueness of submitted structures. These unique numbers were called CNUMBERS and were used for indexing and searching for chemical information. CNUMBERS could be retrieved by performing a substructure search or by searching explicitly for a name or molecular formula. In addition to CNUMBERS, the actual thesaurus indexing terms were used instead of the previous alphanumeric term codes. Searching was based on Boolean logic, as were the earlier computer search systems. A unique feature of the thesaurus was its interaction with the text file and with the document search system. Thus, it was the cornerstone of searching the text file by controlled terms. No term was entered into the search file unless it was in the thesaurus. A search for chemical structure in the second-level CS4 system would link by CNUMBER through the documents to other information about chemicals, such as properties, uses, processing, and so forth. Other search options included the ability to select reports by issue date, document source, or the type of report, for example, a research progress report or a market research report.

In 1966 a study was undertaken to determine the future requirements of the central report group with the goal of designing a system to handle the increasing workloads and new services more efficiently (Montague & Schirmer, 1968). Between 1966 and 1971 a detailed file organization scheme for the Information Flow System was developed for the IBM 360/65 (later replaced by an IBM 370/155) (Hoffman, 1972). Programs were written to convert the IBM 1410/7010 system to the Information Flow System. As in earlier systems the Information Flow System linked retrieval of information about chemical compounds and documents containing information about the compounds via a compound number or CNUMBER (Schultz, 1974). Other important features of the new system included the use of threaded lists in addition to inverted files to optimize searching and a file of report abstracts. Search answers were printed either as a list of accession numbers or abstracts. The printed abstract feature was a welcome replacement for the previous practice of pulling and refiling abstract cards for the screening of search results.

Although the Information Flow System was state-of-the-art for its time, by the early 1980s the twenty-year-old software, though aging gracefully, was becoming unsupportable, owing to system limits, hardware obsolescence, and other factors. Interactive online searching became available in the 1970s for external literature through Chemical Abstracts Service, Dialog, Orbit, and other vendors of secondary information. DuPont scientists and the information scientists in the Central Report Index wanted an online interactive search system for their proprietary report information.

Online Systems

CRIDB. To respond to the needs of the DuPont scientists as quickly as possible, in 1985 an online free-text searchable database (CRIDB) of just abstracts and bibliographic information was implemented. The abstracts and bibliographies of documents in the Information Flow System were copied to a Basis database. (Basis is a product of Information Dimensions, Inc. [IDI].) CRIDB became available in May 1986 and was used until 1991, when it was replaced. CRIDB was searchable by command as well as by menus to accommodate those who were uncomfortable with using commands.

SCION (Scientific Corporate Information Online). In the 1980s Chemical Abstracts Service offered a private registry service. At the time when DuPont investigated this option to replace its aging Information Flow System, several organizations maintained private chemical files with CAS. Since the DuPont chemical file had been developed in conjunction with CAS in the 1970s, it was thought that an electronic conversion could be accomplished easily, although over the intervening years, differences in structure conventions had developed. After reviewing practices for handling text and chemical structures within the chemical industry, corporate information science decided that the private registry service offered the best fit for DuPont's needs. In 1986 discussions were held with CAS on system analysis, design, and development of a chemical search file, a text file for bibliographic information and abstracts, and a hierarchical thesaurus.

During 1987 and through 1988 the DuPont-CAS team designed, built, tested, and implemented the chemical file. All structures and structure-related files for the 160,000 compounds from the Information Flow System were converted to CAS hardware and software in Columbus, Ohio. New chemical input resembled standard registry format. DuPont and CAS had diverged over the years in some structure conventions for chemical classes. The first task was to convert DuPont connection tables to a format that was compatible with CAS input format. About 75 percent were converted electronically; the other 25 percent were converted by manually drawing and then keyboarding the drawn structures.

The document file was designed to parallel as closely as possible the files on STN International Network, especially File CA. Consequently, searching both in-house literature and public literature was possible with one command language and one log-in. The text file was designed to accommodate a field for the links and roles assigned during conceptual analysis. Abstract text and the fields for title, author, document numbers, issue dates, and some other information were converted to a generalized format provided by CAS. Abstracts were converted to allow sentence-level proximity searching.

The Central Report Index thesaurus was converted to the format of existing hierarchical thesauri on STN. It serves three functions: an authority or control list for conceptual analysis vocabulary; a reference for assigning indexing terms or locating terms for searching; and creator of a "generic" or "family" collection of search terms. This third feature is an improvement over the previous practice of up-posting documents to all the terms at the higher levels in the hierarchy at the time of document input. With a generic or family search a collection of search terms based on the broad term–narrow term relationships in the thesaurus is created at search by generating a Boolean logic union or using "or" logic to combine all of the narrow terms with their broad term.

Work on the document file conversion began in 1989 and was completed in 1990. Testing took place during 1990, and the necessary internal DuPont and external telecommunication links for access to the database in Columbus were determined and installed. SCION officially became available to the DuPont community in June 1991 (Marcali, Kvalnes, Patterson, & Wilks, 1993).

To accommodate searchers who did not want to learn STN command language, a DuPont team designed menu screens and the navigation routes between them for menu-assisted searching, which were sent to CAS for implementation. Menu-assisted searching became available in 1992. Work is now under way to convert SCION to an open client-server architecture, which will add additional functionality and broaden its use.

Web-Enabled Information Management

Web-enabled technology began to have its impact about 1992. At that time a company-wide Internet-intranet

Figure 6.
The Intranet

- Corporate news
- Benefit information
- Product specifications
- Electronic commerce
- Technical publications
- Gateway to World Wide Web
- InfoNavigator catalog and contact pages for DuPont businesses
- Extranet

guidance team was established. A general-purpose server was made available to business units for setting up Web sites. There are page counters for statistics, a search engine, forms-based e-mail, password-protected areas, and software-downloading capability (Figure 6).

As Du Pont's intranet continues to grow, it will become the foundation for accessing many types of information and knowledge repositories. The library catalog is part of the intranet, and corporate information science is developing its own suite of home pages to lead DuPont scientists to the multitude of information resources available. Corporate information science also has the responsibility for cataloging sites on the intranet to allow for effective access.

Summary

The factors that have had the greatest influence on the handling of technical information at DuPont are concept coordination; development of a controlled or standard vocabulary for indexing of the large collection of scientific and technical information; atom-bond-atom connection tables, which enable the storage and retrieval of information about specific chemicals; and the computers that made implementation of the other three developments possible. The changing nature of the company's organization has led to a network of libraries and information centers around the world, which provide records management and search support for their local communities of scientists and engineers. These information centers rely on the central unit for purchasing, acquisitions, cataloging, and arranging access to externally published and online resources.

References

Ayers, R. S. (1994–1995). Turning your vision into reality. *Bulletin of the American Society of Information Science, 21(2),* 20–22.

CAS and DuPont to collaborate. (1964, July 20). *Chemical & Engineering News*.

Conrad, C. C. (1955, September). The role of centralized and decentralized information services. Presented at Division of Chemical Literature, American Chemical Society Meeting, Minneapolis, MN.

Costello, J. C. Jr. (1961). Storage and retrieval of chemical research and patent information by links and roles in Du Pont. *American Documentation 12,* 111–120.

Dinwiddie, S. M., & Conrad, C. C. (1954). Report indexing by hand-sorted punched cards. In B. H. Weil (Ed.), *The technical report* (pp. 303–316). New York: Reinhold.

Duncan, V. L. (1951). The Lavoisier library, experimental station, E. I. DuPont de Nemours & Co. *Bulletin of the Special Libraries Council of Philadelphia and Vicinity, 17*(5), 33–37.

Edge, E. B., Fisher, N. G., & Bannister, L. A. (1957). System for indexing research reports using a punched card machine. *American Documentation, 8.*

Gluck, D. J. (1965). A chemical structure storage and search system developed at DuPont. *Journal of Chemical Documentation, 5(1*), 43–51.

Herner, S., Lancaster, F. W., & Johannigsmeirer, W. F. (1964, September 3). Investigation of the effect of roles and links on the performance of a mechanized retrieval system. Presented at Division of Chemical Literature, 148th National Meeting, American Chemical Society, Chicago, IL.

Hoffman, W. S. (1968). An integrated chemical structure storage and search system operating at Du Pont. *Journal of Chemical Documentation, 8 (1),* 3–13.

Hoffman, W. S. (1972). DuPont information flow system. *Journal of Chemical Documentation, 12(2),* 116–124.

Kilgour, F. G. (1997). Origins of coordinate searching. *Journal of the American Society for Information Science, 48(4),* 340–348.

Lancaster, F. W., & Mills, J. (1964). Testing indexes and index language devices: The ASLIB Cranfield project. *American Documentation, 15,* 4.

Luhn, H. P. (1957, October). A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of Research and Development*.

Marcali, J. G., Kvalnes, F. H., Patterson, J. A., & Wilks, E. S. (1993). The Du Pont global technical information system. In W. A. Warr (Ed.), *Chemical Structures 2* (pp. 145–160). Berlin: Springer-Verlag.

Montague, B. A., & Schirmer, R. F. (1968). Du Pont central report index: System design, operation, and performance. *Journal of Chemical Documentation 8*, 33–41.

Morgan, H. L. (1965). The generation of a unique machine description for chemical structures technique developed at Chemical Abstracts Service. *Journal of Chemical Documentation 5,* 107–113.

New system identifies and stores formulas. (1963). *Chemical & Engineering News, 41*(49), 35–36.

Nichols, M. T., Sikes, J., Isselmann, M. M., & Ayers, R. S. (1995–1996). Survival in transition or implementing information science core competencies. *Bulletin of the American Society of Information Science 22(2),* 11–15.

Patterson, J. A., Schultz, J. L., & Wilks, E. S. (1995). Enhanced polymer structure, searching, and retrieval in an interactive database. *Journal of Chemical Information and Computer Science 35*(1), 8–20.

Rasmussen, L. E., & Van Oot, J. G. (1969). Operation of DuPont's central patent index. *Journal of Chemical Documentation, 9*(4), 201–206.

Salton, G. (1961, November). The manipulation of trees in information retrieval. *DDC (Defense Documentation Center) (in NTIS No. AD274816).*

Schultz, J. L. (1974). Handling chemical information in the Du Pont central report index. *Journal of Chemical Documentation, 14*, 171–179.

Schultz, J. L. (1975). Polymer nomenclature, classification, and retrieval in the DuPont central report index. *Journal of Chemical Information and Computer Sciences, 15,* 94–100.

Schultz, J. L., & Wilks, E. S. (1997) Multiple-level polymer registration in the DuPont SCION database. *Journal of Chemical Information and Computer Science 37*(2), 165–170.

Taube, M., et al. (1953). *Studies of coordinate indexing, Vols. 1–1V.* Washington, DC: Documentation Inc.

Taube, M., et al. (1962, February). The state of the art of coordinate indexing. *DDC (Defense Documentation Center) (NTIS No. AD275393).*

Van Oot, J. G., Schultz, J. L., McFarlane, R. M., Kvalnes, F. H., & Riester, A. W. (1966). Links and roles in coordinate indexing and searching: An economic study of their use, and an evaluation of their effect on relevance and recall. *Journal of Chemical Documentation 6,* 95–101.

Wall, E. (1959). A practical system for documenting building research. Presented at Building Research Institute, Washington, DC.

Wilks, E. S. (1997a). Polymer nomenclature and structure: A comparison of systems used by CAS, IUPAC, MDL and DuPont. 1. Regular single-strand organic polymers. *Journal of Chemical Information and Computer Science 37*(2), 171–192.

Wilks, E. S. (1997b). Polymer nomenclature and structure: A comparison of systems used by CAS, IUPAC, MDL and DuPont. 2. Aftertreated (post-treated), alternating/periodic, and block polymers. *Journal of Chemical Information and Computer Science 37*(2), 193–208.

Wilks, E. S. (1997c). Polymer nomenclature and structure: A comparison of systems used by CAS, IUPAC, MDL and DuPont. 3. Comb/graft, cross-linked, and dendritic/hyperconnected/star polymers. *Journal of Chemical Information and Computer Science 37(2*), 209–223.

Wilks, E. S. (1997d). Polymer nomenclature and structure: A comparison of systems used by CAS, IUPAC, MDL and DuPont. 4. Stereochemistry, inorganic, coordination, double-strand, polysiloxanes, oligomers, telomers. *Journal of Chemical Information and Computer Science 37(2*), 224–235.

For Further Reading

Crow, J. E. (1967). Microforms and technical information. Presented at Division of Chemical Literature, American Chemical Society Meeting, Chicago, IL.

Grandine, J. D. 2nd, Starr, E. M., & Putscher, R. (1959, September). Report index searching on the Bendix G-15D computer. Presented at Division of Chemical Literature, American Chemical Society Meeting, Atlantic City, NJ.

Schirmer, R. F. (1967). Thesaurus analysis for updating. *Journal of Chemical Documentation 7,* 94.

Walker, J. F. (1964). The singularity sub-link—a new tool for use in the storage and retrieval of information. *Journal of Chemical Documentation 4,* 45.

Building Information Retrieval Systems for Science

Development of an Information Retrieval System Based on Superimposed Coding

James M. Cretsos

Abstract

This paper describes the design and historical development of an information retrieval system for company-generated scientific and technical reports. The system is based on coordinate indexing, optical coincidence, and tri-axial (three-dimensional) coding and uses descriptors for total recall and descriptive phrasing to answer highly specific queries. A vocabulary with a maximum capacity of one million terms can be manipulated using only 300 optical coincidence (Termatrex) cards. A tri-axial coding scheme is incorporated to reduce the mechanical noise present in densely packed systems to a negligibly low value. A mathematical model predicts first approximations of false drops owing to mechanical noise.

The proliferation of scientific and technical information in the post–World War II precomputer era sparked the development of mechanized and semimechanized information systems for storing and retrieving information. Driven by industry's need to retrieve information contained in internally generated documents, corporations began to establish information facilities to store, retrieve, and disseminate corporate information. Newer information technologies, developed to retrieve information and data from corporate documents, included processing coded media—cards, film, tape—by hand or mechanical devices.

Introduction

While hand-sorted and mechanically sorted card systems had been available decades earlier, their use began to flourish during the 1950s and early 1960s. Because of their low cost, versatility, and easy maintenance, card-based systems gained wide acceptance among individuals and information centers.

Generally, cards were referred to as punched cards and were of two types: edge-notched cards and aspect cards (Casey, Perry, Berry, & Kent, 1958). Although either type could be used to retrieve data or limited amounts of information, the majority of applications involved retrieval of document representations, for example, abstracts, bibliographic citations, or document accession numbers. They became known as the information retrieval systems of that era.

However, most card-based information retrieval systems became unwieldy, as the size of vocabularies or document collections grew. Random-number superimposed coding methods were invented to facilitate searching, although they, too, had their limitations: cumbersome coding for storage and false drops or noise encountered during the retrieval process (Gilbert, 1958).

Superimposed Coding

The simplicity of direct coding (a single alpha or numeric code per single hole) of hand-sorted punched-card systems limited the vocabulary size used to index, search, and retrieve surrogates of documents. The maximum number of index terms a single card could accommodate was equal to the total number of holes on the perimeter of the card. To accommodate more index terms, punched-card manufacturers produced larger cards, thus increasing the number of holes. They also produced cards with two rows of holes on the perimeter of the cards. Even that was not sufficient to alleviate this problem.

In contrast to direct coding to represent a single index term, concept, or descriptor, superimposed coding expanded the capacity of a punched card to accommodate more index terms by compressing several bits of information into a small space. However, superimposed coding often resulted in erroneous selections or false drops. Despite that limitation, a number of card-based systems used superimposed letters or numbers. Two examples of superimposed numbers follow:

1. In combination number coding, the simplest form of superimposed coding, the perimeter of a card was divided into fields of four holes each, with each hole having a printed number, 7, 4, 2, 1, respectively. Any digit from zero (no punch) to 9 could be represented by punching one to two holes in each field. For example, to represent the number 3, holes 1 and 2 were punched (sum of 1 plus 2); to represent the number 6, holes number 2 and 4 were punched, and so on. By using each field to represent a series of numbers, such as units, tens, hundreds, and thousands, one could accommodate a large number of index term representations using only a few holes.

   However, numeric coding contributed to a large number of false drops during card sorting. For example, if an index term was assigned the number 593 and the corresponding holes on the card were punched,

   - 4 and 1 (for 5) in the hundreds field
   - 7 and 2 (for 9) in the tens field
   - 2 and 1 (for 3) in the units fields

   during the card-sorting process, that card could drop erroneously when one was conducting a search also using other terms that had been assigned any of the following numbers: 121, 122, 171, 172, 421, 422, 471, 472.
2. Mooers (1947) applied a different type of numeric coding, using superimposed random number codes for his Zatocoding card system. His Zator cards had designated positions for forty notches, marked 1–40, on the top of each card (eventually, Mooers in 1955 enhanced his system by using both the top and the bottom of the card). Each descriptor was assigned a set of four two-digit random numbers (1–40). Examples of Mooers descriptors and their corresponding Zatocodes are:

   - camera 1 8 29 31
   - flash 17 23 34 38
   - film tally 14 17 22 30

   Random number superimposed coding minimized the probability of getting false drops during the sorting of Zator cards.

Historical Development

The information retrieval system described in this paper was developed at the Technical Information Center of Melpar, Inc., in Falls Church, Virginia (1963 to 1966), as part of an internal research and development program. Jerry Evans and David Vachon, formerly with Melpar, collaborated with the author during the design and development of the tri-axial coding scheme. The system was designed to retrieve documents from a collection of internally generated scientific and technical reports and proposals.

Melpar's primary business was contract services to government and industry, mostly in science, technology, and engineering. To acquire government contracts, Melpar relied heavily on solicited and unsolicited proposals, and it subscribed to the then-popular "shotgun" approach: the more proposals submitted to the government, the greater the chances of winning contracts. Consequently, much effort went into proposal writing, and the company developed outstanding graphics and publications departments to support that effort.

The second salient feature of Melpar's approach to doing business was its universality. The company bid on virtually everything in which the U.S. government was interested—covering the gamut of science, technology, and engineering as well as education, social sciences, housing, economics, and so on. Senior scientists and engineers at Melpar spent most of their time in their offices writing proposals. Because of the usually short turnaround time to respond to requests for proposals, writers were continually pressed to meet deadlines. To speed up the writing process, scientists and engineers would often selectively lift relevant information and data from previously submitted company proposals and technical reports. Having fast access to company reports and proposals became essential reference aids to writing new proposals.

Need for Information Storage and Retrieval

Melpar estimated that it had about 38,000 scientific and technical reports and proposals in its internal document collection. Documents varied in length—anywhere from twenty pages for a mathematics research report to several thousand pages for a series of technical reports on lunar expedition modules. Because many of those documents were classified, all reports and proposals were stored in secure file cabinets and administered by the company's security department. Users could request documents, on a "need-to-know" basis, by proposal or report number, the name of the contracting agency, agency document number, or government document number.

As the document collection grew in size, retrieval of individual reports and proposals became an unwieldy and time-consuming process. Melpar management recognized the dire need for a better system to retrieve documents and turned to its Technical Information Center for help.

Melpar management's mandate to the Technical Information Center was clear and direct: "Design and develop an affordable information retrieval system that is easily searchable by scientists and engineers." That meant a card-based system and an open-ended vocabulary. It also meant learning more about the information searching and gathering habits of Melpar's scientists and engineers.

System Design and Development

Nature of the Melpar Document Collection

To get a sense of the nature of the internal document collection, our staff examined three hundred technical reports and proposals, selected at random from each of the major disciplines. It became obvious from the nature and complexity of internal documents and the highly diverse needs of Melpar's scientific and engineering personnel that there was a need to design an information retrieval system that would allow coordination of index terms and accommodate an open-ended vocabulary.

Information-Seeking Behaviors

A team of three information specialists held structured interviews with two hundred scientists and engineers to learn about their information-seeking behaviors. Data gathered from those interviews revealed that the information-seeking and -gathering habits of scientists and engineers differed along disciplinary lines. Engineers looked for tables of data, drawings and designs, and manufacturing processes. Biologists and medical personnel sought information about causative agents and their effects on plants, animals, and the environment as well as prevention and treatment. Astronomers and physicists were interested in theories, tools for massaging numbers, and physical properties of materials or conditions. Mathematicians and economists were concerned more with mathematical models and proofs. Behavioral and social scientists looked for reasons and statistical explanations of how certain conditions and situations affect individuals and communities. Chemists, especially organic chemists, were most demanding, seeking specific information on methods, chemical reactions, derivative compounds, catalysts, and thermodynamic data.

Review of Other Systems

We reviewed two systems based on the concept of links and roles. The first was developed by Costello (1961) at DuPont for a system of over 200,000 documents and a vocabulary of 137,000 terms. Links and roles were used to help prevent false coordination of terms. The second system, developed by Logue (1962) at Monsanto, also used links and roles and a word list of 28,000 terms to index about 23,000 technical reports. While the use of links and roles was suitable for indexing many of Melpar's chemical reports, it could not have been applied to most of the scientific and engineering reports in the collection.

Mechanized Systems under Consideration

Of the various punched-card systems used to retrieve documents, my collaborators and I selected two for investigation: the Zatocoding system, developed by Mooers (1955) and the Termatrex system, developed by Jonker (1960). Each system had its own proponents, and both had been purchased by major corporations to help retrieve internal company technical reports.

Zatocoding

The Zatocoding system was designed to handle special edge-notched cards and could accommodate a limited vocabulary up to 350 descriptors and a collection up to 25,000 documents. It was based on coordinate indexing and superimposed coding of four sets of two-digit random numbers, for coding descriptors, and "scatter coding" of letters, for coding names. Even with its Zator 200 Selector equipment, the system was limited to handling two hundred cards at one time. Searching a collection of 25,000 documents required 125 sets of needle sorting, and sorting time depended largely on the digital dexterity of the operator.

A low-cost, constrained set of descriptors that enabled indexers and searchers to think in terms of broad concepts, random card filing, and random number superimposed coding were the key advantages of the Zatocoding system. Time-consuming searching and false drops—retrieval of irrelevant cards—were its major disadvantages. In addition Garfield (1961) addressed a problem inherent in superimposed coding systems: the failure to consider the importance of term utilization, that is, how many times a term was used in the coding and retrieval process.

Termatrex

The Termatrex system belonged to the aspect class of punched-card systems, also known as inverted punched-card systems, superimposable card systems, optical coincidence systems, or more commonly peek-a-boo systems. Aspect card systems differed significantly from edge-notched card systems, in design, physical attributes, and mode of searching. Instead of one card representing a single document with its associated terms coded on the peripheral edges of the card, each card represented a single term. Holes in the body of the card represented document accession numbers.

The Termatrex system was based on coordinate indexing of terms and comprised sets of a thousand cards, divided into ten groups of a hundred cards per group. The cards in each group were color coded and tabbed numerically from 00–99 for easy access. Each Termatrex card represented one index term and had a maximum capacity of ten thousand holes or ten thousand document accession numbers.

To code a document with its associated terms, all cards corresponding to the index terms of the document were selected and superimposed. Then a hole was drilled through all the cards simultaneously. Each hole represented a serial number on an x–y matrix of 100×100 positions. The cards could be filed randomly within each color group. Using several sets of Termatrex cards, the system could accommodate up to two hundred thousand documents and several thousand terms.

To retrieve a set of documents that contained information and data related to a search query, Termatrex cards representing corresponding terms were selected, superimposed, and placed atop a light box. Light coming through the holes of superimposed cards corresponded to accession numbers of documents that contained the information and data relevant to the search query.

Major advantages of the Termatrex system were its relatively low cost, constrained random filing, flexibility, ease of use, and convenience. However, the Termatrex system lost many of its advantages as the number of terms or the number of documents in the collection increased. For each additional one thousand terms or ten thousand documents, a second set of one thousand Termatrex cards was required. For example, ten thousand Termatrex cards would have been required to handle a collection of ten thousand documents using ten thousand terms, or one million Termatrex cards to handle a collection of one hundred thousand documents using one hundred thousand terms.

Nature of Vocabulary

The aim of our research project was to design a system that could guide Melpar users to find information and data contained in the company's internal documents. Its scope was to enable users to search the document collection on broad subject areas or highly specific topics, a system that was designed to retrieve, for example, all documents dealing with antennas, as well as all documents containing information on the construction of airborne C-band antennas.

Ordinarily, the method used to coordinate individual terms, *construction* and *airborne* and *C-band* and *antennas*, also could have led the user to erroneous documents related to *airborne construction of C-band antennas*, meaning the construction of C-band antennas during flight. Stringing terms to form descriptive phrases eliminated false coordination of terms. The indexer or searcher would have to ask two types of questions:

1. What is it? To identify the subject.
2. What kind or type is it? To identify each of its modifiers.

For example, in the case of *airborne C-band antennas*, one would ask:

What is it?

It is an *antenna* (or *antennas*).

What kind of an antenna?

It is an *airborne antenna*.

What kind of an airborne antenna?

It is a *C-band airborne antenna*.

The string phrase, *C-band airborne antenna*, became a single phrasal term.

Selection of Terms for Indexing and Searching

Indexers and searchers used different approaches to selecting terms. To index the topic *C-band airborne antenna*, an indexer would select *antenna, airborne antenna, C-band airborne antenna*. To search for the same topic, a user needed to select only one term: *C-band airborne antenna*.

If the user wanted to broaden the search to find documents with information on *airborne antennas*, the searcher would select the phrasal term, *airborne antenna*. Similarly, by selecting the term *antenna*, the searcher would retrieve the documents containing information on all types of antennas.

This process of prelinking or precoordinating terms addressed the problem of false coordination of terms as exemplified by the *venetian blind–blind Venetian* problem. For example, to search for information on *blind persons from Venice*, one could ask,

What is it?

It is a *Venetian* (a person from Venice).

What kind of Venetian?

It is a *blind Venetian* (a blind person from Venice).

Similarly, to search for *venetian blinds*, one would ask,

What is it?

It is *blinds* (could be wooden, aluminum, venetian, vertical, etc.).

What kind of blinds?

It is *venetian blinds*.

Bernier and Crane (1962) posed a similar problem related to *aluminum-molybdenum* and *molybdenum-aluminum alloys*. The former is an aluminum-rich molybdenum alloy; the latter, a molybdenum-rich aluminum alloy. Use of uniterms, *aluminum*, *molybdenum*, and *alloy* would retrieve information on both aluminum-rich molybdenum alloys and molybdenum-rich aluminum alloys. However, one could achieve specificity by asking,

What is it?

It is an *alloy*.

What kind of an alloy?

It is a *molybdenum alloy*.

What kind of molybdenum alloy?

It is a *molybdenum-aluminum alloy*.

Tri-Axial Coding

We recognized at the outset that to achieve high specificity in retrieval would require an expansive vocabulary of tens of thousands of terms. The Termatrex system of one card for each term could accommodate that requirement. However, with a collection of nearly forty thousand documents, the system would become too unwieldy, if one had to handle more than two hundred thousand Termatrex cards.

While random number superimposed coding of terms had been used at that time solely on edge-notched card systems, such as Zatocoding, we considered applying the concept of superimposed coding to Termatrex cards.

In the Termatrex system of one card per term, each term also could be represented numerically from 00–99 as the x axis (top, horizontal edge) of the card. With the color-coding and numeric indexing properties of Termatrex cards, one could represent a term as a color-coded number, like, Red 35, Green 02, Blue 79, or Yellow 93, and so on. Instead of terms having linear numeric designations, our approach was to assign each term a position within the three-dimensional space of a cube, such position described by its x, y, and z axes.

In our view a tri-axial coding scheme (Figure 1) of 100 × 100 × 100 could accommodate a large vocabulary of up to one million terms using only three hundred Termatrex cards, one hundred cards for each axis. A table of one million random numbers was used to select and assign a single but unique six-digit number to each term. The six-digit random number was separated into three groups of two-digit numbers, one two-digit number for each of the x, y, and z axes of the cube. The color and numbered-tab properties of Termatrex cards made it convenient to assign each unique six-digit random number to three cards of different colors, with red representing the x axis, blue the y axis, and green the z axis. For example, the six-digit random number code assigned to the term *Device(s)*, 82-71-72 would be applied to three different Termatrex cards: Red-82, Blue-71, and Green-72.

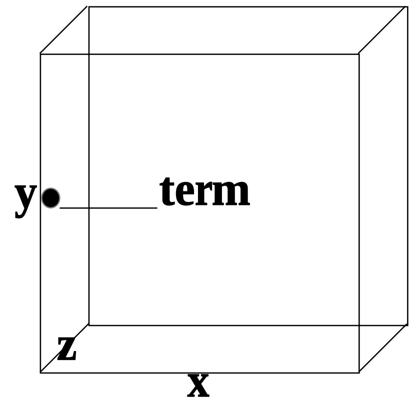


Figure 1. Tri-axial coding.

Probable false drops, that is, retrieval of erroneous documents, an inherent property of information retrieval systems based on superimposed random number codes, were of some concern. In the Zatocoding system, for example, it was not uncommon to get a few false drops per sort of two hundred cards. To predict the probable frequency of false drops in our system, we applied the following mathematical expression:

$$E_1 = 1 - \exp\left(\sum_{d \leq D} \log_e f(d)\left\{1 - \exp\left[d\sum_{k=1}^{n}\left(\log_e(c_k - a) - \log_e c_k\right)\right]\right\}\right)$$

E = probability of obtaining at least one erroneous document

d = number of terms used to index a document

D = total number of documents

c_k = length of k^{th} axis

a = total number of terms in vocabulary

Table 1. Probability of False Drops When One Term Is Used to Index Two Documents

Maximum Number of Documents Indexed by Each Term	Total Number of Terms in Vocabulary	Total Number of Documents	Maximum Number of Allowable False Drops
2	1,000	1,000	0.51
		3,000	0.98
		5,000	1
		7,000	1
		10,000	1
	5,000	1,000	0.43
		3,000	0.94
		5,000	1
		7,000	1
		10,000	1
	10,000	1,000	0.36
		3,000	0.92
		5,000	1
		7,000	1
		10,000	1
	50,000	1,000	0.36
		3,000	0.89
		5,000	0.99
		7,000	1
		10,000	1
	100,000	1,000	0.35
		3,000	0.87
		5,000	0.97
		7,000	1
		10,000	1
	500,000	1,000	0.35
		3,000	0.72
		5,000	0.87
		7,000	0.94
		10,000	0.98
	1 million	1,000	0.35
		3,000	0.72
		5,000	0.87
		7,000	0.94
		10,000	0.98

Table 1 shows the maximum number of false drops for infrequently used terms, as the size of vocabulary and size of the collection grow.

When a term is used to index more than five documents, the number of false drops becomes negligible, as shown in Table 2.

System Prototype

Equipment

Equipment for Melpar's information retrieval system included a Jonker 202 Manual Data Input Device (for drilling Termatrex cards) and a Jonker 52A Reader (light source for reading superimposed Termatrex cards). The Termatrex system included a filing bin with a thousand Termatrex cards and a set of transparent color cards (yellow, blue, red, and green).

Pilot Study

We selected a thousand technical reports and proposals representing eight scientific and technical disciplines for a pilot study. A committee representing Melpar's scientific and technical community and the Technical Information Center made the selection. Documents varied in size and technical complexity.

Document Processing

A team of three information specialists with backgrounds in biology, chemistry, engineering, mathematics, and physics reviewed, abstracted, indexed, and coded each document using the tri-axial coding scheme. Documents were indexed in depth, with an average of 47.3 terms per document. Indexers were also involved in the search and retrieval process. A member of the clerical staff assigned document accession numbers, drilled the Termatrex cards, and maintained a coded vocabulary on Rolodex cards.

Vocabulary Development

The Melpar vocabulary of index terms consisted of a set of descriptors and a list of phrasal terms under each descriptor, as shown in Figure 2. To index 1,000 documents for this pilot study, indexers generated a vocabulary of 8,300 descriptors and phrasal terms.

The Armed Services Technical Information Agency and the Engineers Joint Council thesauri (1960, 1964) were used to develop descriptors. In indexing documents,

Table 2. Probability of False Drops When One Term Is Used to Index Five or More Documents

Maximum Number of Documents Indexed by Each Term	Total Number of Terms in Vocabulary	Total Number of Documents	Maximum Number of Allowable False Drops
5	1,000–1 million	1,000	$< 1 \times 10^{-8}$
	1,000–1 million	5,000	$< 1 \times 10^{-8}$
	1,000–1 million	7,000	$< 1 \times 10^{-8}$
10	1,000–1 million	10,000	$< 1 \times 10^{-6}$
	1,000–1 million	1,000–10,000	$< 1 \times 10^{-20}$
> 25	1,000–1 million	1,000–10,000	$< 1 \times 10^{-50}$

DEVICE	82-71-72	
Generalized–		17-02-28
digital–		24-90-64
Homing–		06-88-39
Intercommunication–		00-39-75
Instrument–		83-02-21
multi-engine–		64-21-31

Figure 2. Vocabulary card.

information specialists followed guidelines for creating and selecting descriptors and phrasal terms. The guidelines were designed to help avoid the problems related to viewpoint, generics, and semantics as outlined by Holm and Rasmussen (1961).

Testing

The system was tested using three hundred queries submitted by the authors of technical reports and proposals and by naive users seeking information. Information specialists handled the searching or guided the users to selecting descriptors and phrasal terms. Retrieval of documents was, for the most part, on target. Two documents were retrieved in error, and one document was not retrieved.

Analysis

A postmortem examination of the system's three failures showed that the causes were owing to human error. Retrieval of two erroneous documents was caused by faulty transposition of numbers by the clerical staff during the term coding process. Failure to retrieve one document was caused by indexer error. The document was a proposal dealing with the deployment of bombardier beetles as harassing agents during military conflict. While the information specialist had indexed the document under *bombardier beetles*, he had failed to index it under the name of the genus and species of the bombardier beetle, a name that was mentioned only once in a document of more than 150 pages.

Evaluation

Cretsos, Evans, and Vachon (1965) reported that the results of post-test interviews with fifty users indicated 98 percent satisfaction with the system. The sole dissatisfied user felt documents were not indexed deeply enough to meet his information needs. Users also made several suggestions to improve the display of terms and expressed a strong need to have printed personal copies of the vocabulary.

References

Bernier, C. L., & Crane, E. J. (1962). Correlative indexes. VIII. Subject-indexing vs. word-indexing. *Journal of Chemical Documentation, 2,* 117–122.

Casey, R. S., Perry, J. W., Berry, M.M., & Kent, A. (1958). *Punched cards: Their applications to science and industry* (2nd ed.). New York: Reinhold Publishing Corporation.

Costello, C. J. (1961). Storage and retrieval of chemical research and patent information by links and roles in DuPont. *American Documentation, 12,* 111–120.

Cretsos, J. M., Evans, J. W., & Vachon, P. D. (1965, April 7). Evaluation of semantic and coding designs of an information retrieval system based on superimposed coding. *Division of Chemical Literature, 149th National Meeting of the American Chemical Society.* Detroit.

Garfield, E. (1961). Information theory and other quantitative factors in code design for document card systems. *Journal of Chemical Documentation, 1,* 70–75.

Gilbert, P. T. (1958). An optimal punch-card code for general files. *American Documentation, 9,* 84–98.

Holm, B. E., & Rasmussen, L. E. (1961). Development of a technical thesaurus. *American Documentation, 12,* 184–190.

Jonker, F. (1960). The Termatrex inverted "punched-card" systems: History of their development, their operation and use. *American Documentation, 11,* 305–315.

Logue, P. (1962). Deep indexing technical reports. *Journal of Chemical Documentation, 2,* 215–219.

Mooers, C. N. (1947). Putting probability to work in coding punched cards—Zatocoding. *Division of Chemical Education, 112th Meeting of the American Chemical Society* (15–19 September 1947). New York.

Mooers, C. N. (1955). Zatocoding and developments in information retrieval. *Aslib Proceedings, 8,* 3–22.

Thesaurus of ASTIA descriptors. (1960). Washington, DC: Armed Services Technical Information Agency.

Thesaurus of engineering and scientific terms. (1964). New York: Engineers Joint Council.

The Evolution of Citation Indexing—From Computer Printout to the *Web of Science*

Jacqueline Trolley and Jill O'Neill

Abstract

Citation indexing was developed in the late 1950s as a new way to monitor, organize, and retrieve the literature. The *Science Citation Index* was one of the first large-scale, machine-generated indexing systems. Over the course of forty years it has become an essential tool for the scientific community. In particular, the *SCI* provided a new dimension in indexing, permitting the researcher to trace the literature both retroactively and retrospectively. Thus the *SCI* complemented traditional bibliographic databases which are designed to assist the researcher with current awareness, to aid in retrieving relevant material from an ever-larger body of literature, and to help separate the more relevant from the mass relevant publications.

Citation indexing was conceived in the early 1950s as a way to monitor, organize, and retrieve published scientific and scholarly literature. While citation indexing had been implicit in legal citators such as *Shepard's Citations,* the concept had not yet been applied to the literature of any field of scientific research. *Shepard's Citations* came into existence in 1873 to provide the legal profession with a tool for tracking subsequent decisions based on cases decided by federal and state courts (Adair, 1955). The *Science Citation Index* (*SCI*), launched by the Institute for Scientific Information (ISI) in the early 1960s, was one of the first applications of computers in the production of large-scale, machine-generated indexes.

SCI's development was intimately related to the earlier implementation of *Current Contents* (*CC*). In the abstract sense the *SCI* could have been created de novo. In practical terms it was the availability of the collection of current journal issues that made it possible to proceed with large-scale experimentation. *CC* (launched as *Contents in Advance* in 1953) was designed to help scientists become aware of what was being published in core journals central to their own investigations as well as in peripheral journals. It started as a customized service to drug firms in 1957, but by 1958 it was published as *Current Contents/Chemical, Pharmaco-medical and Life Sciences* (Garfield, 1993). It later was expanded into the physical, social, clinical, engineering, agricultural, and arts and humanities editions. Over the course of the late 1950s *CC* demonstrated a multidisciplinary approach to the scientific community.

The Beginnings

The Johns Hopkins Medical Indexing Project was sponsored by the Army Medical Library, later the Armed Forces Medical Library, which ultimately became the National Library of Medicine. Located at the Johns Hopkins University Welch Medical Library in Baltimore, it was established in 1948 to investigate the role of automation in the organization and retrieval of medical literature. In addition to studying machine methods of compiling indexes, the project investigated the human process of selecting subject headings, descriptors, or other indexing terms. The goal was to reduce the human element and thereby increase the speed of cataloging current biomedical articles and including the index entries into the published indexes.

Prompted by a suggestion from Chauncey D. Leake, then chairman of the advisory committee to the Welch project, Eugene Garfield, a member of the project team, investigated the nature and linguistic character of review articles and how they dealt with the literature reviewed. Garfield recognized that they indirectly "indexed" each of the many papers cited. Each sentence in the review, which identified an original published source for a notable idea or concept, was an indexing statement (Garfield, 1993). By capturing these references and organizing them into an inverted list, the researcher could get a view of the approach taken by another scientist to support an idea or methodology based on the sources consulted and cited. Thus, the addresses of the papers—bibliographic citations—could be assigned by a professional indexer.

In designing the scope of the putative citation index and its cost-effectiveness, Garfield was aware of Bradford's "law of scattering" (Bradford, 1953). Bradford had observed that in any given field of investigation a relatively small group of journals represented the core of the field. However, Garfield discovered that the essential core of journal literature for all fields of scientific research is found in a basic group of five hundred to a thousand journals. Different sets of journals from this basic core will have a greater relevance to one topic and lesser relevance to others. Garfield used the analogy of a comet, "the nucleus representing the core journals of a literature and the debris and gas molecules of the tail of the comet representing additional journals that sometimes publish material relevant to the subject" to describe this (Garfield, 1979). Garfield also observed that the tail of the literature of one discipline consists, in large part, of the core of the literature of another discipline; this is now referred to as Garfield's "law of concentration" (Garfield, 1979). Thus as a complement to Bradford's law, Garfield applied his law of concentration and found that by monitoring the core journal literature in several scientific fields one could optimize the cost-effectiveness of the database. Core journals included those that produced not only the most articles but also the most influential papers as measured by the frequency of citation. However, the Welch Project was terminated in June 1953 before these ideas could be tested.

Early Years at ISI

After the Welch Project, Garfield attended Columbia Library School and began a career as a documentation consultant. He formed DocuMation, Inc., in 1955, which initially published *Management's Documentation Preview*. It ultimately became *Current Contents Management* in 1956, when DocuMation became Eugene Garfield Associates. This firm became the Institute for Scientific Information in 1960.

Eugene Garfield Associates conducted two pilot projects to test the viability of citation indexing. The first involved the creation of a database based on the references cited in five thousand chemical patents held by Merck and other companies. Garfield's collaborator at Merck, Marge Courain, had been a fellow graduate student at Columbia. The references cited in this database were mainly to prior patents, the documentation sources used by patent examiners to support a decision to grant or deny a patent claim. Garfield and Courain compared the retrieval connections that their experimental patent citation index permitted with those obtained using the Patent Office's classification system. They found that citation indexing retrieved relevant patents that were missed by the Patent Office's current classification system (Garfield, 1979). Eugene Garfield Associates also worked on indexing chemical compounds for the Patent Office under contract with the Pharmaceutical Manufacturers Association.

Several years later, the second, much larger project was launched. In 1960 a grant was obtained from the National Institutes of Health (NIH) and the National Science Foundation (NSF) to build and test a citation index to the published genetics literature. Three test databases were to be created to cover the literature spanning one year, five years, and fourteen years. Each database would include material from a varying number of source publications. The five- and fourteen-year indexes would test the reliability of a narrow, traditional, discipline-oriented *Genetics Citation Index*. The one-year index would test a broader multidisciplinary genetics index, including the emerging field of molecular biology. This index would cover a broadly based set of source publications, since genetics had been invigorated by Watson and Crick's discovery of the DNA double helix in 1953. The emergent field of molecular genetics involved subjects as diverse as crystallography, biochemistry, genetics, and physics. Indeed, some of the early relevant papers in molecular biology were published in the *Review of Modern Physics*. The one-year database drew not only on journals in the field of traditional genetics research but also on a large hardcore interdisciplinary pool of journals ancillary to genetics and molecular biology.

The project employed the automated IBM punched-card system, but workers were still required to key and standardize the varied citation formats. However, the project demonstrated the overall cost-effectiveness of machine-based citation indexing in comparison with traditional human subject indexing. While recognizing the value of natural language title indexing, adopted early on in *CC*, the prime basic objective of the project was to produce the *Genetics Citation Index* proper. The *Genetics Citation Index* would permit the user to determine whether and where any article or book was cited.

During the life of the *Genetics Citation Index* Project the NIH changed its policy of providing grants to all types of organizations. The new policy required that contracts be negotiated with for-profit organizations. Consequently, the NSF was given the task of administering the contract, which also later included a study of coverage by traditional abstracting services. That study demonstrated the many gaps in article coverage of many journals, especially those with multidisciplinary scope. In particular the letters and other editorial items were often missed. So the *SCI* later adopted a full coverage policy.

At the project's completion the government sponsors chose not to subsidize the development of an ongoing citation index database. Garfield made the financially risky decision to move ahead with the private publication of the already prepared multidisciplinary index. The first edition of the *SCI*, covering 1961 source literature, required six volumes. It was made available for purchase in 1963. The *Permuterm Index* was added to *SCI* as an outgrowth of experience with producing weekly subject indexes for *CC* (Garfield, 1957). Permuterm indexing was designed in 1964 by Garfield and his research collaborator, Irving Sher. It involved pairings of words from article titles in such a way as to give users with limited information a way into the multidisciplinary coverage of *SCI*, even if they did not have a particular paper in mind. It eliminated some of the difficulties associated with keyword-in-context (KWIC) indexes (or rotated indexes), which were popular at the time. Permuterm indexing required that all titles be in English, necessitating translation.

In 1964 the *SCI* was launched on a current quarterly basis with an annual cumulation. In 1970 a five-year cumulation covering 1965 to 1969 was produced. Eventually cumulated citation indexes for 1945 to 1954 and 1955 to 1964 were created. Few people, even at ISI, believed that the costs of these indexes could be recovered, but Garfield believed the leading research libraries of the world would eventually buy these indexes for historical and sociological research. He felt they were essential to the future value of *SCI* as a tool for contemporary history of science and technology. His prediction proved correct.

The *Social Sciences Citation Index* (*SSCI*) was launched in 1965 and its source literature now goes back to 1956. The *Arts and Humanities Citation Index* (*A&HCI*) was started in 1975. Since 1980 the *SCI*, *SSCI*, and *A&HCI* have been offered in CD-ROM format. Also in 1975 ISI's new *Journal Citation Reports* (*JCR*) was included as the last volume of the *SCI*. The *JCR* would eventually become a separate service. *JCR*'s current impact factors and other citation data have a great influence on journal and research evaluation worldwide (Garfield, 1976).

Standard measures of relevance made popular by the Cranfield group led by Cyril Cleverdon could not be applied to the evaluation of citation indexing because by using cited reference searching a researcher was, in fact, able to retrieve papers that at first glance might not seem relevant to his or her study. Yet these references often proved crucial to research, and users of the *SCI* soon recognized this advantage. The *SCI*, *SSCI*, and *A&HCI* are considered today to be among the most reliable resources for tracing the development of scientific or scholarly ideas beginning with the primordial papers or books on any given topic.

In 1997 ISI launched a Web-based and completely integrated continuation of the *SCI*, *SSCI*, and *A&HCI*. Known as the *Web of Science*, it bridges the cultures of the arts and sciences, providing integrated coverage of all the academic disciplines via the Internet or intranets. Citation networks are an inherently hypertext approach to navigation of the literature: Users can instantaneously search the bibliographic literature independent of time. Bibliographic coupling, called related records, provides an additional method of clustering documents.

It is significant to the history of citation indexing that Garfield began his career as a chemist. In the same period that *CC* was growing from one to seven editions across the academic and industrial spectrum, Garfield also pursued his dream of a unique chemical information service. Thus, in 1960 he launched the *Index Chemicus*. It is now approaching its fortieth anniversary and has culminated in the development of an integrated chemical compound and reaction database fully linked to the citation index.

This short review of the history of ISI and its work in developing citation indexes has omitted numerous details. Many of these have been reported on in a thoughtful investigation by Paul Wouters in his remarkable doctoral dissertation (Wouters, 1999).

References

Adair, W. C. (1955). Citation indexes for science. *American Documentation 6*, 31.

Bradford S. C. (1953). *Documentation* (2nd ed.). Washington, DC: Public Affairs Press.

Cleverdon, C. (1970). Evaluation tests of information retrieval systems. *Journal of Documentation 26*(1), 55–67.

Garfield, E. (1957). Breaking the subject index barrier—a citation index for chemical patents. *Journal of the Patent Office Society,34,* 583–589.

Garfield E. (1976). The Permuterm Subject Index: An autobiographical review. *Journal of the American Society for Information Science, 27,* 288–291.

Garfield E. (1979). *Citation indexing: Its theory and application in science, technology and humanities.* New York: John Wiley & Sons, Inc.

Garfield E. (1993). *Essays of an information scientist: Of Nobel class, women in science, citation classics and other essays* (Vol. 15). Philadelphia: ISI Press.

Wouters, P. (1999). *The citation culture.* Doctoral dissertation, Science and Technology Dynamics, University of Amsterdam, The Netherlands.

The Creation of the *Science Citation Index*

Paul Wouters

NEWS RELEASE — Contact: Mrs. Joan E. Shook

INSTITUTE FOR SCIENTIFIC INFORMATION 33 SOUTH SEVENTEEN STREET PHILADELPHIA, PA phone/locust4-4400 cable/currcon twx/ph 803

For Immediate Release

$300,000 GRANT TO PROBE INFORMATION RETRIEVAL AWARDED TO INSTITUTE FOR SCIENTIFIC INFORMATION BY NATIONAL INSTITUTES OF HEALTH AND NATIONAL SCIENCE FOUNDATION . . .

THREE YEAR PROJECT TACKLES CITATION INDEX TECHNIQUES FOR SCIENCE

Research scientists will soon be consulting a more precise and specific literature index that links together subject material that would never be collated by usual indexing systems. Concerned with new starting points for scientific literature searches, the unique concept uncovers sometime-buried associations, relating important works and authors, yet keeps the researcher abreast of the masses of current published scientific information. This new approach to information retrieval is called the Citation Index.

A $300,000 grant extending over a three-year period has been awarded to the Institute for Scientific Information, Philadelphia, Pennsylvania, to study the practicability of citation indexes and to test their techniques of preparation. The project, under joint sponsorship of the National Institutes of Health and the National Science Foundation, is aimed at producing a unified citation index for science, including the publication of a genetics index.

Two years after the above press release, the *Genetics Citation Index* was published (Garfield & Sher, 1963). It was quickly followed by the first volume of the proper *Science Citation Index (SCI)* (Garfield, 1963). Since then, the *SCI* has become part of the world of science. This is not to say that the *SCI* was immediately applauded. Initial responses were mixed (Wouters, 1999). These mixed feelings were nevertheless far more positive than the reactions Eugene Garfield, the inventor of the *SCI*, had received in the preceding years. He had been actively propagating the idea of a citation index for science since he became acquainted with it in 1953. Hardly anyone had responded. Even a few years before getting the decisive grant from the National Institutes of Health (NIH) and the National Science Foundation (NSF), referees of his proposal were quite critical, some even hostile. An undated and anonymous overview of the referees comments can be found in Eugene Garfield's personal archive in Philadelphia, Pennsylvania.

This resistance to the idea of investing in citation indexes of the scientific literature may come as a surprise to the present-day user of the *SCI* and *Social Science Citation Index*. After all, scientists must acknowledge their peers and must share their ideas and resources with their colleagues. Therefore, it seems rather obvious to use the footnotes of a scientific article or the bibliography of a book as an entry in a literature-searching procedure. The fact of the matter is, however, that the concept of the citation index did not come to science as naturally as this interpretation suggests. To start with, the *SCI* has its roots not in science but in law.

Stumbling over the Citation Concept

Citation indexes were already old hat for American lawyers when the history of *SCI* begins. In the second half of the nineteenth century Frank Shepard in Illinois deemed it useful to know whether a legal case was still valid. He produced gummed paper with lists of the cases that cited the case at hand. Lawyers in Illinois glued them to their dossiers so enthusiastically that Shepard set up a commercial business in 1873. His company, Shepard's Citations, Inc., had the monopoly of producing the one and only citation index "to serve the Bench and Bar." First in Chicago, later in New York, and in the 1950s in Colorado Springs, Colorado, a staff of highly qualified lawyers produced the Shepard's Citator by hand, covering all judicial decisions in the United States. Shepard's was a respectable firm, proud of its supreme reliability. Its product was grounded in the norms and procedures of the legal system. Shepard's was purchased in 1996 by Reed Elsevier and the Times Mirror Company. As William C. Adair, former vice president of the company, explained to the readers of *American Documentation* in 1955:

> The lawyer briefing a case must cite authorities to back up his arguments. So must the court in writing its opinions. This is because of the doctrine of "Stare Decisis," which means that all courts must follow precedents laid down by higher courts and each court generally also follows its own precedents. . . . The lawyer, however, must make sure that his authorities are still good law, that is, that the case has not been overruled, reversed, limited or distinguished in some way that makes it no longer useful as a valid authority. Here is where the use of Shepard's Citations comes in. (Adair, 1955)

The searching procedure was simple. First, the lawyer located a case similar to his own, then he looked up Shepard's to see whether later cases had cited it. He would immediately see whether the decision was still valid and which other cases had made use of it. A lowercase "r" before the case meant that it was reversed. Adair told his audience that important law suits were won "on the strength of a case located by the use of Shepard which no other method of research disclosed" (Adair, 1955). "Shepardizing" the legal literature was based, since 1873, on the authority-centered norms in the United States legal system: The most recent decision of the highest court is valid. The way of indexing by citation tied in perfectly with this value system. One can hardly think of a sharper contrast with supposedly ruthless scientific criticism. This hierarchical indexing style served nevertheless as the model for the Institute of Scientific Information's *Science Citation Index*.

Retired, running a cattle ranch in Colorado Springs, and still eager to act, William Adair sometime in 1953 read in the local newspaper that the scientific world "was being swamped in a sea of literature." It was a report on a conference organized by the Welch Medical Indexing Project at Johns Hopkins University in Baltimore, Maryland. This project had been sponsored by the Army Medical Library since 1948 (Larkey, 1949; Miller, 1961). The main task of the project was to find out whether, and if so how, machines could be used to improve the efficiency of indexing and retrieving medical literature. The indexing itself was supposed to be the good old subject indexing form. In this respect the Welch Medical Library was not very innovative. Within these boundaries, the staff had to devise new systems of indexing, subject headings, and using machines to solve "the literature problem." Adair wrote a letter to Sanford Larkey, supervisor of the project. He told Larkey about the citation indexing system, informing him that "if the whole body of American Law can be classified so that a knowledge of one case can be used as a key to locate all other cases in point, the same thing can be done with medical articles" (W. C. Adair to S. V. Larkey, personal communication, 10 March 1953). Adair offered his expertise: "I have retired from Shepard's and am now free to undertake and organize such a project." He received a reply from a 25-year-old junior member of the staff—Eugene Garfield—who did not know anything about citation indexing. He wrote Adair that his suggestion would be investigated but kept him at a distance. "We do not have any positions open for staff members," Adair was told (E. Garfield to W. C. Adair, personal communication, 16 March 1953). Nothing happened. Adair's initiative had no impact on the Medical Indexing Project.

More than a year later, after he had been fired by Larkey, Garfield (Eugene Garfield, personal interviews, 27 January 1992 and 4 February 1992) resumed contact with Adair "with the idea of writing a paper to be published in one of the learned society journals" (E. Garfield to W. C. Adair, personal communication, 11 June 1954). Having browsed through *Shepard's Citations* at the public library, Garfield was intrigued with the idea and had even written a paper on "Shepardizing the Scientific Literature" while he was a student at Columbia University (Garfield, 1954). At first, Garfield was not certain whether citation indexes could be applied to science:

> Without knowing exactly what you had in mind I do not feel it is fair for me to be discouraging at the outset. But the one thing that must be kept in mind when comparing the field of science with that of law, is that there are anywhere from one to three million articles each year appearing in the scientific journals. (E. Garfield to W. C. Adair, personal communication, 11 June 1954)

Garfield did not yet think of building a citation index. Working as a consultant in automation, he mainly focused on possible uses of computers (E. Garfield to W. C. Adair, personal communication, 11 June 1954). He perceived an opportunity in automating the production of citation indexes. Garfield and Adair decided to write two papers on the subject (Adair, 1955; Garfield, 1955). Through their correspondence, Garfield learned about the way Shepard's produced its citator, which familiarized him with the ins and outs of citation indexing.

The Citation Introduced to Science

Garfield's article was published in *Science*:

> In this paper I propose a bibliographic system for science literature that can eliminate the uncritical citation of fraudulent, incomplete, or obsolete data by making it possible for the conscientious scholar to be aware of criticisms of earlier papers. It is too much to expect a research worker to spend an inordinate amount of time searching for the bibliographic descendants of antecedent papers. It would not be excessive to demand that the thorough scholar check all papers that have cited or criticized such paper, if they could be located quickly. The citation index makes this check practicable. (Garfield, 1955, p. 108)

The index would be very handy for the working scientist: "It is best described as an association-of-ideas index, and it gives the reader as much leeway as he requires." In this respect the citation index would be, Garfield stressed, far superior to the traditional subject indexes that by nature restrict the interpretation of the article to a predefined number of topics (Garfield, 1955). Not only did Garfield focus on the information needs of the scientist, but he also translated the concept of the citation index in terms of the subject indexes with which both scientists and librarians were more familiar. His and Adair's articles did not, however, attract much attention.

Garfield, now an independent documentation consultant who advised, among others, Smith, Kline & French, was not deterred by the silence that followed his proposal. He undertook several initiatives to make the citation index more popular, which increased his grip on the intellectual and practical difficulties in compiling such an index. Together with Margaret Courain, supervisor of the Research Files Division at Merck, Garfield produced an experimental citation index to patents that he presented at the Minneapolis meeting of the American Chemical Society on 16 September 1955 (Garfield, 1957). At the December 1955 meeting of the American Association for the Advancement of Science in Atlanta, he made a strong plea for a centralized national documentation center (Garfield & Hayne, 1955). Partly as a personal exercise Garfield prepared a citation index to the Old Testament that he presented in 1956 to the American Documentation Institute in Philadelphia (Garfield, 1956). In this talk Garfield presented a new idea, interpretative citation indexing.

In January 1957 Garfield received the first serious support from a scientist. "Dear Mr. Garfield," wrote geneticist Gordon Allen, then at the Department of Health, Education, and Welfare of the NIH:

> Since the appearance of your article in *Science* two years ago, I have been eagerly looking for some news of steps toward a citation index. I have urged the American Society of Human Genetics to take some initiative in the matter [Allen had done this in 1956, when he talked to Sheldon Reed, then president of the society, (G. Allen, personal communication, 9 April 1959)], but they are already involved in the construction of a subject index in human genetics. The references I have seen to your suggestion (for want of a citation index, I probably have not seen all of them) have been disappointingly cool, and I wonder if you have received any personal letters that were

> more enthusiastic. If a group of interested persons were brought together, they might be able to make some headway. (G. Allen to E. Garfield, personal communication, 24 January 1957)

Stimulated by this support, Garfield submitted a proposal to NSF (Garfield, 1958) in August 1957. Its goal was "to determine the utility of citation indexes for science in terms of general usefulness, invariance in time, minimizing the citation of poor data, identification of the 'impact factor,' and provision for individual clipping services." The study also should develop "a suitable technical design for citation indexes." Its motivation followed Garfield's line of reasoning, albeit with an added emphasis on the index's potential for "the encyclopedic integration of scientific statements." Garfield also emphasized the potential of John Desmond Bernal's 1948 proposal for a central clearinghouse. The project was meant to be a two-year study, starting September 1958, and would restrict itself to compiling the index. One month later NSF turned down Garfield's proposal but expressed interest in a citation index (Dwight Gray to E. Garfield, personal communication, 23 October 1958).

Garfield took this as a flat refusal, proving once again in his mind NSF's inability to deal with the tasks at hand. He did not stop his campaign, though. In November he made his plea for a "unified index to science" at the National Academy of Science's Conference on Scientific Information (Garfield, 1959). In this presentation the idea of integrating scientific knowledge, already mentioned in his 1958 article, was further developed (Garfield, 1959, p. 674).

In May 1959 Garfield received a letter from geneticist Joshua Lederberg, which would prove the turning point in the history of the *SCI*:

> Since you first published your scheme for a "citation index" in *Science* about 4 years ago, I have been thinking very seriously about it, and must admit I am completely sold. In the nature of my work I have to spend a fair amount of effort in reading the literature of collateral fields and it is infuriating how often I have been stumped in trying to update a topic, where your scheme would have been just the solution! I am sure your critics have simply not grasped the idea, & especially the point that the author must learn to cooperate by his own choice of citations + thus he does the critical work. Have you tried to set this out in an adequate experiment? Would you look for support from the NSF? Of course you have to count on opposition from the established outfits, which have already succeeded in blocking any progressive centralization of the Augean tasks. (J. Lederberg to E. Garfield, personal communication, 9 May 1959)

As Lederberg later explained (29 July 1960) to Garfield, Lederberg's initiative was prompted by a science policy debate in the Genetics Study Section of NIH. The administration wished to evaluate its actual impact on research and proposed, in the words of Lederberg (J. Lederberg to E. Garfield, personal communication, 29 July 1960) "a number of rather fancy and inefficient schemes." Lederberg recognized that a citation index would accomplish the purpose "at a negligible additional cost" and decided to contact Garfield. Garfield's reaction was enthusiastic:

> I hope you won't be embarrassed by a show of emotion, but your memo almost brought tears to my eyes. It then seemed that over six years of trying to sell the idea of citation indexes had not been completely in vain. (E. Garfield to J. Lederberg, personal communication, 21 May 1959)

He told Lederberg the whole story of his pleas for citation indexes, the support of Gordon Allen, and the resistance he had met since 1954 (E. Garfield to J. Lederberg, personal communication, 21 May 1959). Lederberg was shocked by Garfield's letter and replied that he was "absolutely astonished that citation indexes are not long since a standard feature at the Patent Office" (J. Lederberg to E. Garfield, personal communication, 18 June 1959). He advised Garfield in the same letter to resubmit his proposal "to all the agencies who could be interested." The NIH he said, "would be an excellent target," since it "is anxious to evaluate its 'impact' on scientific progress, and how better do this than through your scheme." Lederberg proposed Garfield to "jump in" and ask NSF's assistance in organizing a scientific committee as suggested by NSF's last letter on the subject.

This resulted in the proposal to NSF (Garfield, 1960a) to construct a citation index by scanning a list of a thousand journals and processing all references to forty-three specified genetics journals as well as all references to twenty-two specified general science journals. The proposal to NIH, on the contrary, entailed the processing of "references from the specified journals, punching them into IBM cards, and mechanically sorting these. Only then would the citations to genetics journals be selected and printed" (Garfield, 1960b). The index would be published by turning over to the editors of the journals "an individual journal citation index" that they could publish as a yearly supplement. This was only an "intermediate mechanism," though. Garfield wished to keep open the option of a separate publication and wrote that he was "in correspondence with editors on the publication problem." On 26 December 1960 he could at last break the big news (the "notification and statement of grant award" granting $49,450 per year for three years) to Lederberg: "Dear Josh, the official note that NIH approved our grant came in the other day. This was quite nice Xmas present to say the least" (Brewer to E. Garfield, 15 December 1960, received December 23, 1960; E. Garfield to J. Lederberg, personal communication, 26 December 1960).

And to Allen: "Dear Gordon, Santa Claus was very good to us. We learned that NIH approved its half of the revised budget which NSF asked me to submit based on $100,000 per year for three years" (E. Garfield to G. Allen, personal communication, 26 December 1960). The NSF grant was approved two months later.

Building the Index

"I think you're making history, Gene!" wrote Lederberg on 24 January 1962. Building the *SCI* turned out to be a bigger project than even Garfield expected. It took more time, more money, and was technically more complicated than foreseen in the contracts. Constructing the index was not only a technical endeavor, but a political enterprise as well. Joshua Lederberg perceived the *SCI* as a means to open up the clogged communication channels in science. Building the index required not only a technical or library expertise but political acumen as well. Through the intense cooperation of Garfield and Lederberg, this enterprise became part of the science policy debate in the United States on the now-famous Weinberg report (PSAC, 1963). The history of the *SCI* can hardly be understood without an appreciation of these technical difficulties and political dimensions.

Science policy provided the context in which Lederberg remembered Garfield's 1955 proposal and decided to write him. *SCI*'s political relevance was directly related to its bibliographic properties. Yet the connection between the *SCI* and science policy was rather loose. From October 1961 onward this changed. The building of the *SCI* became intimately involved in the debate on the future of scientific information in the United States after Joshua Lederberg was appointed a member of the Panel on Science Information (PSAC). His assignment was to rewrite the general introduction to the Weinberg report, and he saw this as an opportunity to push for a radical overhaul of the "anarchic" way scientific information was organized. Lederberg advocated a centralized information system, modeled after John Desmond Bernal's pleas from 1948. In the end Lederberg argued, in a letter to Garfield on 8 October 1961, that this should result in the abolition of the traditional journals.

Lederberg took this political development as an opportunity also to promote the *SCI* itself and asked Garfield to be his "informal consultant" and give him background information on "detailed proposals that you consider reasonably intelligent." Garfield showed no hesitation (E. Garfield to J. Lederberg, personal communication, 11 October 1961).

The PSAC issued its report in 1963 with an array of proposals and calls for action, directed to the federal government, to the scientific community, to individual scientists, and to the libraries. PSAC (1963) has since been seen as a landmark in the history of documentation (Schneiders, 1982, p. 176). In this report the information crisis was not merely a question of keeping the individual scientist informed, as it had been formulated in the 1958 Baker report (PSAC, 1958). The crisis threatened the very identity of science. The panel opened the report with the following sweeping statement:

> Science and technology can flourish only if each scientist interacts with his colleagues and his predecessors, and only if every branch of science interacts with other branches of science; in this sense science must remain unified if it is to remain effective. The ideas and data that are the substance of science and technology are embodied in the literature; only if the literature remains a unity can science itself be unified and viable. Yet, because of the tremendous growth of the literature, there is danger of science fragmenting into a mass of repetitious findings, or worse, into conflicting specialties that are not recognized as being mutually inconsistent. This is the essence of the "crisis" in scientific and technical information. (PSAC, 1963, p. 7)

The PSAC report called for drastic action and for major changes in the scientific system and the behavior of individual researchers. One of the recommendations was the development of a new searching tool, the citation index, about which the panel was "particularly impressed."

These recommendations were, at least partly, the result of an intense correspondence between Lederberg and Garfield about the solution to the problem of scientific information while they were building the *SCI*. Garfield had laid out a comprehensive scheme comprising "three levels of reporting: title, abstract, full paper." The basic idea was that 1 percent of the papers would be published in a "national or international organ" (for example, a daily science newspaper); the next 10 to 25 percent would be published in "a series of select journals," whereas the vast majority of the papers would be put in a central depository. This would put an end to the proliferation of new journal titles. The newspaper would also publish lists of all papers (E. Garfield to J. Lederberg, personal communication, 10 November 1961). The national documentation center, which would be the central axis, should distribute "a series of abstract journals." Moreover, a "prompt translation service" would provide for fast international communication.

Garfield envisioned his *Current Contents* or its successor as the place to "publish by title," whereas the newspaper would have a daily citation index section. All in all the system would be a drastic improvement for timely access to all available information. "An important factor," Garfield stressed in a 10 November 1961 letter to Lederberg, "is that a man's personal bibliography should have the same publication value regardless. A reference to a paper that does not get into the major primary organ or in the journals should be considered equally." Garfield reiterated Bernal's ideas in a letter to Lederberg, dated 15 November 1961, as he made clear by urging Lederberg to look into Bernal's papers.

In the first draft of his "Notes on a Technical Information System," Lederberg reconstructed the main problem as follows:

> As members of the scientific community we have a deeply rooted obligation to interact with the "literature." Not so much the size but the dispersion and formlessness of the institution make this an ever more hopeless aspiration. . . . The present system has generated two responses: the defeat of neurotic frustration for some, the compromise of narrow specialism for others. I feel the survivorship of humanistic science demands a better solution. (Lederberg, 1962a, p. 1)

Lederberg proposed a central depository together with "select journals":

> A centralized repository would provide the range of materials that I would specify as being required for my immediate and retrospective information requirements. Concurrently, select journals with high standards of selection and editorial quality would maintain my contact with the breadth of scientific culture. (Lederberg, 1962a, p. 7)

The depository would be built according to a set of ground rules. One of these would be that no paper could be withdrawn once deposited: "As with journal publications the author's reputation is permanently attached to it." Papers would be distributed and refereed "promptly." Moreover, an updated citation index would be attached to the articles. The principal advantage of the repository scheme was, according to Lederberg, the "prompt and widespread availability" of contemporary findings. "That contributions can take a full year to come out in print is an absurdity of modern science" (Lederberg,1962a, p. 4). The repository would "discourage the redundancy implicit in peripheral publication and in the irresponsibility of gossip and 'invisible colleges.' " It would facilitate the publication of "expensive archival documents" like taxonomies. Last but not least, it would stimulate the journals to "revert to being *select* journals: they are broadsides on which I would rely to bring me *unasked* the best or overtly most interesting of contemporary science." The user would be more central than in the prevailing system, Lederberg felt. He expected the "journal output" to decrease to "about 10% of its current level."

The central problem with realizing this radical overhaul was that it needed a certain critical mass. Hence, the idea of a daily newspaper, which Garfield and Lederberg had discussed in their conversations about the *SCI*, also became a strategic item in realizing their information revolution. The already existing publication system, with its vested interests, seemed the main obstacle (E. Garfield to J. Lederberg, personal communication, 10 November 1961). Unfortunately, Lederberg noted, "one of the serious shortcomings of the *OSIS* in NSF is that it really has neither the staff nor the mandate to consider such large scale systems propositions." Garfield had the same experience. He had sent his proposal for a unified index to science in newspaper format to NSF but received no response (E. Garfield to J. Lederberg, personal communication, 6 March 1962).

Lederberg's scheme differed fundamentally from conventional publication in scientific journals. First of all, the primary responsibility for seeking editorial criticism would be shifted to the author. Second, the need for primary journals would disappear. "Relieved of the unnatural responsibility for primary archives and communication," the scientific societies and other journal sponsors could devote themselves to too-often neglected services "especially in review and interpretation." At this stage of scientific communication Lederberg wished more opportunities for commercial initiatives (Lederberg, 1962b, p. 4). Third, authors would also be responsible for the production of abstracts, since "manpower requirements" prevented their central production. Lederberg acknowledged the possibility that "peripheral agencies" also might be able to continue their abstracting services. Fourth, the government would have the primary responsibility for financing the whole system. Fifth, the system would be oriented to innovation, looking to "the future development of data handling and telecommunication systems to replace the techniques of the present proposal."

These policy discussions stimulated Garfield to see the Institute for Scientific Information's (ISI's) role in terms of shaping the future of scientific communication:

> I have been thinking "big" down here in terms of ISI's future. I hope to incorporate this thinking into a series of proposals that tie in with your proposals on Science Advisory Committee. . . . I am convinced we are only five to ten years away from bridging the existing artificial gap between technical science writing and writing for the laymen. In fact, there is probably a greater need than you and I realize for a citation index "structure" that would relate a conventional clipping service with our scientific clipping service. (E. Garfield to J. Lederberg, personal communication, 9 July 1962)

This view tied in with the problem of how to publish the *SCI*. While the computer programs, data files, and citation indexing procedures were being developed, the question of publishing the resulting index became more pertinent. Garfield proposed to Ralph O'dette in a 17 September 1962 letter that NSF test "the newspaper format" for a daily citation index, to achieve a "low cost per reading." The newspaper should have the format of the *New York Times*, initially comprise sixteen pages, and contain reprints of original research papers and review articles (four pages), a daily updated author bibliography (five pages), a citation index (six pages), and a subject index (one page) (Garfield, 1962). The author bibliography would contain 750 papers per day and was vital for the use of the indexes. Garfield expected that in one year three million citations would have been listed this way. The "Daily Scientist" as it was called should be a throw-away paper: "The philosophy behind a daily dissemination technique is that the information comes in small segments. The daily newspaper is quickly scanned and then discarded" (Garfield, 1962, p. 2). Garfield estimated that scientists would be prepared to pay a subscription fee of thirty dollars a year. He proposed that NSF test the idea by sending 25,000 scientists consecutive daily issues for two months. If the NSF would give initial support, the experiment could be expanded with the help of NIH, NASA, and AEC. A one-year experiment would cost around $500,000, Garfield estimated. Most of this money would be necessary to produce a unified citation index to science anyway. Therefore, Garfield argued, his proposal would "bring a vast amount of information to the individual scientist at a phenomenally low cost" (Garfield, 1962, p. 2).

The NSF, however, was not prepared to fund the production of the *SCI*, which resulted in Garfield's decision to publish the index on a for-profit basis. The risky adventure nearly bankrupted ISI, and it was mainly on the profits generated by other products, primarily *Current Contents*, that the *SCI* had a chance to become profitable.

Conclusions: Translating the Citation Concept

Automation

The citation index NIH and NSF supported and the *SCI*, as it would be published from 1964 onward, did not look like *Shepard's Citations*. Technically, the idea was still the same. Because of this, Garfield's proposal to NSF could state that most of the uses of the *SCI* were "analogous to their use in legal research." This statement nevertheless concealed essential dissimilarities. The fundamental change was in the meaning of a citing relation. The outlook of the index differed as well. Moreover, ISI's way of producing the index would be the complete reversal of Shepard's. The production of the *SCI* was therefore not a matter of simply applying a ready-made tool in a novel area. Developing the *Science Citation Index* required both a new way of looking at the scientific literature and a new conception of citation indexing: "The brilliant utility of the citation index approach is that it cuts across the problem of meaning by an automated procedure" (J. Lederberg to E. Garfield, personal communication, 9 November 1962).

It would have been impossible to make a database such as the *SCI* without computers because it would have been far too expensive. Even with existing computers, it was a risky business. It was computerized processing that made possible the migration of the citation concept from the legal to the scientific context. The corresponding devaluation of labor made the production of the *SCI* possible within the budgets available for these kinds of enterprises in the United States at the time.

Comprehensiveness

By automating the production of the *SCI*, Garfield, Lederberg, and Allen could tackle the enormous task of indexing the scientific literature while retaining its complete coverage of science. William Adair had been aware of the problems of scale as well but did not think of automation. Instead, he proposed to index separately the various scientific specialties or disciplines. It was a familiar solution; *Shepard's Citation* also was fragmented according to the structure of the legal system in the United States. The idea was not strange in the world of science either. After all, most scientific journals were limited to a narrowly defined specialty. Moreover, several other citation index projects were constructed along the same lines. Garfield had been the principal propagandist for citation indexing, but he was not the only one involved. In the early 1960s NSF supported several citation index research projects. The *SCI* project was, however, the only attempt to produce a comprehensive citation index covering, in principle, all of science.

Two principal "competitors" had opted explicitly in favor of a monodisciplinary approach. Statistician and leading citation index researcher John Tukey studied and built a citation index of the statistics literature at Princeton University (Tukey, 1962; Tukey, n.d. b; Tukey, n.d. a). At the Massachusetts Institute of Technology, the inventor of bibliographic coupling, Michael Kessler (1961), was constructing a complete information system of the physics literature. He did not consider a citation index strong enough to sustain a pilot model system in itself, although it would be a useful element to add once the model was constructed, because citation was "a low probability event" (Kessler & Heart, 1962; Kessler, 1965).

The objective of the *SCI* to cover all the scientific literature was underpinned by "the unity of science." Without the possibility of going beyond the boundaries of the academic disciplines, a citation index would add practically nothing to traditional subject indexes. After all, the researcher could be relied upon to know the literature in his or her own specialty. The *SCI* should be able to locate relevant research in unexpected places, and this only seemed possible if the *SCI* was not structured along disciplinary lines. The *SCI* was also expected to change the citing behavior of the scientist, which was not the case with Shepard's Citator, only a registering device. The citing behavior of attorneys and judges was fairly standardized, which made it possible for indexers to classify citations with a fairly restricted set of symbols. In contrast the reason a scientist cites an article is not restricted at all. In fact references to scientific papers play divergent roles. Even the same citation can change meaning in the course of time. The makers of the *SCI* expected to exert a positive influence on the scientists' citing behavior. In its turn this would increase the value of the *SCI*.

The Information Crisis

In the 1950s science had been growing too fast to cope with its results. It made some parts of the scientific community gradually receptive to innovations in handling the literature. This "information crisis" is a key factor in the birth of the *Science Citation Index*, playing social as well as cognitive roles. It shaped the way the central problems in the realm of science, science management, and science policy were defined. Government agencies provided funds to find solutions to this information crisis and thereby created a new labor market for people with both scientific and librarian skills. This new field was where people as diverse as a documentation specialist, a researcher in human genetics at the National Institutes of Health, a Nobel laureate in bacterial genetics, and a retired vice president of Shepard's could meet each other. The crisis, made more urgent by the *Sputnik* surprise, eventually gave citation indexing the official approval it needed to take off.

It was a debate at NIH about the evaluation of NIH-funded research that reminded Lederberg of Garfield's 1955 paper in *Science* and prompted him to write his memo in 1959. Once a citation score is transformed into a measure of the impact of a paper, all sorts of policy-related studies can be easily imagined if the database is large enough. The sociological use of the *SCI* was an outgrowth of this capability and of the network approach. Notwithstanding, the central motive for scientists like Lederberg and Allen was, and would continue to be, the literature-searching capabilities.

Innovative Outsiders

Without the drive, perseverance, social capacities, and technical expertise of Eugene Garfield, the immense task of building the *SCI* would probably not even have been thinkable. It is not only a matter of personal traits, but also of being in the right place at the right time. Garfield was an outsider in more than one respect, which made it possible for him to think about solutions other people would reject immediately. Garfield was well prepared for information services, *Current Contents* being the proof of that. Not coincidentally, the two scientists who reacted to Garfield's 1955 article in *Science* were geneticists. The structure of the new science of genetics made coping with the literature more pressing for Lederberg and Allen than for, say, the nuclear physicists. Genetics still had unclear boundaries. On the other hand, the professional societies in human and bacterial genetics stuck to the old subject indexing, as did all relevant institutions. Thus the personal histories of Lederberg and Allen have been important factors as well.

In the process of translating the citation concept to the world of science, the funding agencies and Eugene Garfield learned to cope with each other. Garfield was an outsider to the academic world, which does not mean he did not have many contacts with researchers and science policy officials. On the contrary networking was one of Garfield's strong points. He was asked to review proposals to NSF on indexing projects on a regular basis (E. Garfield, personal communication, 12 September 1959). He was running his own company, Eugene Garfield Associates, with *Current Contents* as its main product. He was not affiliated with a respectable academic institution, which created additional hurdles.

An intellectual problem existed as well: Citation indexing was an unknown entity. Garfield's proposals showed this, and naturally he wanted to keep open as many options as possible. But the funding agencies were also uncertain and wanted to know more precisely what they were supposed to support. Allen's and Lederberg's support made Garfield's undertaking more respectable. Moreover, they taught Garfield how to deal with agencies like the NSF and NIH, while transforming the citation concept in this process.

Success as well as Failure, and Yet a Success

The experimental genetics citation index appeared in 1963, the *SCI* in 1964. Since then the *SCI* and its associated products are a well-known feature in most scientific libraries over the world. ISI almost went bankrupt because of the *SCI*, but in the end it turned out to be profitable (E. Garfield, personal interviews, 27 January 1992 and 4 February 1992, Philadelphia). It seems a classic American success story, with log cabin (Garfield's chicken coop in New Jersey where he started producing *Current Contents*) and all. And a success the *SCI* surely is.

But it is also a story of failure. Lederberg was not only thinking about a bibliographic tool when he pushed the case of citation indexing in the courts of science policy, but he also set out to revolutionize the whole publication system of science. In 1959 Lederberg had adopted Bernal's program of doing away with all scientific journals as a primary channel of publication. As a member of the PSAC panel on scientific information, he was impressed by Derek Price's book *Science since Babylon* (1961) and pressed for abolishing the anarchical way of publishing. All commercial publishers should be pushed out of the business of primary publication. The process of scientific communication should be made "efficient, systematic, anxietyfree, reliable." Papers would be available on request, and their existence would be announced via abstract services. The refereeing system would be completely eliminated, authors being responsible for their own products. Retrieval of literature would be rationalized with machine-driven indexes, citation indexing being one of them. A daily journal of science would be the central medium of mass communication in the whole system.

With Garfield acting as his informal consultant on the matter, Lederberg advanced these innovative—and perhaps radical—ideas. In their hands the *SCI* would not be merely a searching tool but a revolutionizing instrument, profoundly changing the world of science. In this respect their enterprise was a failure. The birth of the *SCI* did not make any immediate changes in the scientific community, nor did it profoundly influence scientists' behavior. By limiting the scope of the *SCI*, the existing institutions successfully defended the traditional way of publishing.

And yet on a more fundamental level the *SCI* is a success, but in a different way than its creators expected. While the *SCI* did not trigger immediate changes in the scientific system, it did shape a whole new set of signs of science. The citation indexing concept that Garfield and Lederberg had forged from its original legal citator predecessor became the cornerstone of a novel social science specialty—scientometrics—as well as the building block of an intricate maze of science and technology indicators. As I have argued elsewhere (Wouters, 1999), this development has created a set of fundamentally novel representations of science and technology that has influenced both science policy and the production of scientific knowledge at all levels.

References

Adair, W. C. (1955). Citation indexes for scientific literature? *American Documentation 6*, 31–32.

Garfield, E. (1954). *Shepardizing the scientific literature.* Unpublished manuscript, Columbia University Library School, New York.

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science 122*, 108–111.

Garfield, E. (1956). Citation indexes to the Old Testament. Unpublished.

Garfield, E. (1957). Breaking the subject index barrier—a citation index for chemical patents. *Journal of the Patent Office Society 39*, 583–595.

Garfield, E. (1958). *A general feasibility study of citation indexes for science: A proposal for research.* Philadelphia: Eugene Garfield Associates.

Garfield, E. (1959). A unified index to science. *Proceedings of the International Conference on Scientific Information, Washington D.C., November 16–21, 1958,* (pp. 461–74). Washington, D.C.: National Academy of Sciences–National Research Council.

Garfield, E. (1960a). Citation index to genetics and general science literature: research proposal. Philadelphia: Institute for Scientific Information.

Garfield, E. (1960b). Citation index to genetics literature: Application for research grant. Philadelphia: Personal Archive Eugene Garfield.

Garfield, E. (1962). Proposal: Application of newspaper publishing methods to the dissemination of scientific information. Philadelphia: Institute for Scientific Information.

Garfield, E. (1963). *Science Citation Index: An international interdisciplinary index to the literature of science* (Vol. 1: aabe–capl). Philadelphia: Institute for Scientific Information.

Garfield, E., & Hayne, R. L. (1955, December). Needed—a national science intelligence & documentation center. Paper presented at the American Association for the Advancement of Science, Atlanta, GA.

Garfield, E., & Sher, I. H. (1963). *Genetics citation index—experimental citation indexes to genetics with special emphasis on human genetics*. Philadelphia: Institute for Scientific Information.

Kessler, M. M. (1961). An experimental study of bibliographic coupling between technical papers. Preprint.

Kessler, M. M. (1965). The MIT technical information project. *Physics Today,* 28–36.

Kessler, M. M., & Heart, F. E. (1962). Concerning the probability that a given paper will be cited. Supported by a grant from NSF.

Larkey, S. V. (1949). The Army medical library research project at the Welch Medical Library. *Bulletin of the Medical Library Association 37*, 121–124.

Lederberg, J. (1962a). Notes on a technical information system. Philadelphia: Personal Archive Eugene Garfield.

Lederberg, J. (1962b). SCITEL: A central scientific communication system. Philadelphia: Personal Archive Eugene Garfield.

Miller, E. (1961). The National Library of Medicine index mechanization project. *Bulletin of the Medical Library Association 49*(1), 1–96.

Price, D. J. De S. (1961). *Science since Babylon.* New Haven, CT: Yale University Press.

PSAC (1958). *Improving the availability of scientific and technical information in the United States: A report of the President's Science Advisory Committee*. Washington, DC: The White House.

PSAC (1963). *Science, government and information: A report of the President's Science Advisory Committee*. Washington, DC: The White House.

Schneiders, P. (1982). *De bibliotheek—en documentatiebeweging 1880—1914; bibliografische ondernemingen rond 1900.* Doctoral dissertation, University of Amsterdam, The Netherlands.

Tukey, J. W. (1962). Keeping research in contact with the literature: Citation indices and beyond. *Journal of Chemical Documentation 2*, 34–37.

Tukey, J. W. (n.d. a). *The citation index and the information problem: Opportunities and research in progress.* Princeton, N.J.: Statistical Techniques Research Group, Princeton University.

Tukey, J. W. (n.d. b). *A citation index for statistics: An introduction and a progress report.* Philadelphia: Personal Archive Eugene Garfield.

Wouters, P. (1999). *The citation culture*. Doctoral dissertation, Science and Technology Dynamics, University of Amsterdam, The Netherlands.

Information Retrieval in Science

The Professional Aspects

Wilhelm Ostwald, the “Brücke” (Bridge), and Connections to Other Bibliographic Activities at the Beginning of the Twentieth Century

Thomas Hapke

Abstract

This paper gives a summary of the activities of the German chemist and Nobel laureate Wilhelm Ostwald (1853–1932) in the area of scholarly information, communication, and publication at the beginning of the twentieth century. In 1911 Ostwald, with others, founded the “Brücke” (Bridge), an organization with aims similar to those of the famous Institut International de Bibliographie in Brussels. The paper looks at connections to other institutions and individuals in the area of documentation and “information science,” especially in Germany, for example, the Institut für Techno-Bibliographie and the German librarian Julius Hanauer, one of the German promoters of the Universal Decimal Classification.

Introduction

On 29 November 1915 Morris L. Cooke, a prominent follower of Frederick W. Taylor, who had died a few months before, wrote a letter to the German chemist and Nobel laureate Wilhelm Ostwald:

> Our organization [Frederick Taylor Co-operators, Chestnut Hill, Philadelphia] is planning, so far as at all possible, to take Mr. Taylor’s place in promulgating in this country and abroad matters, which will interest scientific management men wherever they may be. I feel that we know all too little about your work about “Die Brücke,” and any other line of activity of the same general character.” (Berlin-Brandenburgische Akademie der Wissenschaften [BBAW] Ostwald papers, No. 3470)

The last sentence remains true today for Wilhelm Ostwald in the history of scientific information and communication. In 1911 Ostwald, with others, had founded the “Brücke” (Bridge), an organization with aims similar to those of the famous Institut International de Bibliographie in Brussels. Although he also published his ideas and work on the organization of scientific publications in English journals (Ostwald, 1913; Ostwald, 1914) and was mentioned in some contemporary papers (for example by Homer, 1912; Bugge, 1925) and bibliographies (Schneider, 1923), Ostwald remained an outsider to the areas of librarianship and documentation in Germany and abroad.

In the last quarter of this century one finds a few papers published in English, which perhaps change this picture slightly (Holt, 1977; Bonitz, 1980; Satoh, 1987), but much research remains. For a recent German paper, see Hapke (1997). Ostwald’s activities in the area of scientific organization gained more recognition in East Germany, the former German Democratic Republic, for example, Lewandrowski (1979) and Bonitz (1979). Based on research on selected Ostwald papers found in the Akademie-Archiv of the Berlin-Brandenburgische Akademie der Wissenschaften as well as on Ostwald’s published works, this paper summarizes the activities of Wilhelm Ostwald in the area of scholarly information, communication, and publication at the beginning of the twentieth century. In addition, this paper shows some of the connections of Ostwald and the Bridge to institutions or individuals in the bibliographic movement at the beginning of this century to make clear that there was a “bridge,” however small, between the Bridge and others (Satoh, 1987, p. 18).

On the one hand, Ostwald's work was a product of the situation in Germany at the turn of the century (Johnson, 1990); on the other hand, it was related to numerous worldwide movements before World War I: internationalism (Lyons, 1963; Crawford, 1992); energetics (Hakfoort, 1992); taylorism (Burchardt, 1977); encyclopedism, which culminated afterward in the ideas of H. G. Wells (1938); and last but not least, the library and documentation movement (Schneiders, 1982). Today you find a rebirth of some of his ideas together with the developments of hypermedia systems and the World Wide Web (Rayward, 1994, 1997; Buckland & Plaunt, 1997).

Wilhelm Ostwald, born in 1853 in Riga, Latvia, was one of the founders and organizers of physical chemistry at the end of the nineteenth century. In recognition of his role in the chemical profession, in 1887, he was appointed to the only chair of physical chemistry in Germany, at Leipzig; the other candidates withdrew in favor of Ostwald. On the basis of thermodynamics and positivism, he developed his energetics (Leegwater, 1986), which he extended to his philosophy of nature (Naturphilosophie). His so-called "energetic imperative," "Do not waste energy, but convert it into a more useful form" (Holt, 1970, p. 388) was an important foundation for his later efforts with regard to the organization of scholarly work. He resigned from his chair in Leipzig in 1906 to devote more time to philosophy and monism as well as to the international organization of scientific work and to the development of his color theory. In 1909 he received the Nobel Prize in chemistry in recognition of his work on catalysis and for his investigations into the fundamental principles governing equilibria and rates of reaction. Ostwald died in Leipzig in 1932. For a more detailed biography see Rodnyj and Solov'ev (1977) and also Hiebert and Körber (1978) and Fleck (1993). For Ostwald's influence on the history of physical chemistry see Servos (1990). In Ostwald's autobiography (1926–27) he mentioned a number of his organizational efforts in scientific work.

Ostwald's ideas about how science works seem to be modern in one sense (see, e.g., Krohn & Küppers, 1989). He said, for example, on the occasion of the opening ceremony of Jacques Loeb's biological laboratory in Berkeley, "Science is an organism which strives constantly for self preservation and development. It is therefore provided with organs of regulation, by which that which is useful is preserved and that which is harmful suppressed" (Ostwald, 1903b, p. 19, English original). In another sense his ideas of the sciences as a pyramid-shaped building with "Kulturwissenschaft," his name for sociology, on the top, one subject standing on the foundations of the one below (Ostwald, 1929), were typical of his time, representative of positivism and scientism.

Both his views on science and his activities in scientific publication formed the foundation for Ostwald's efforts to organize scientific publication and communication.

Ostwald's Activities in Scientific Publication

Textbooks

In his *Lehrbuch der allgemeinen Chemie* (1885–87), the first textbook on physical chemistry, Ostwald succeeded in reviewing the state of the art and collecting the scattered papers on the subject of physical chemistry, to which little attention had been paid. Later Ostwald stressed the advantages of combining reading such an encyclopedic compilation with browsing in original sources to find new problems (Ostwald, 1903a, pp. 13–14). His ability to follow this regimen was one reason for his great success as a scholar in Leipzig, where he founded his own research school.

Establishing Scientific Journals

Consistent with his emphasis on the original sources of scientific work, in 1887 Ostwald and the Dutch chemist Jacobus H. van't Hoff founded the *Zeitschrift für physikalische Chemie,* the first periodical in physical chemistry (Hapke, 1990; Pohle, 1998). In his view, the problems of publishing physical chemistry papers in "normal" chemical periodicals as well as the difficulty of gaining widespread dissemination of such papers were now solved. From the beginning the enterprise was a fully commercial periodical based on international collaboration. As such, the title page of the first volume lists many collaborators from abroad. The *Zeitschrift* attracted all scientists interested in physical chemistry, who found in the journal not only original papers but also reviews and abstracts of other important works in physical chemistry, both books and papers, from other journals.

In 1894 Ostwald was also engaged in founding the *Zeitschrift für Elektrochemie*. In the new century he was the editor of many more periodicals in his new areas of interest, Naturphilosophie, monism and color theory: the *Annalen der Naturphilosophie* (1.1901/02–11.1912/13; 14.1919/21), *Das monistische Jahrhundert: Zeitschrift für wissenschaftliche Weltanschauung und Kulturpolitik*

(1.1912/13–4.1915), and *Die Farbe: Sammelschrift für alle Zweige der Farbkunde* (1.1921–44.1926).

Publication of the Klassiker

Since 1889 Ostwald had been editing his *Klassiker der exakten Wissenschaften*, original scientific works republished for easy access as separate volumes (Dunsch, 1989). He wanted to counterbalance the growing quantity of journal literature with his selection of papers of lasting importance. In his autobiography Ostwald said that the editing of the *Klassiker* was the "germ for the much later ideas on the technical organization of science" (Ostwald, 1926–27, p. 56). With the same concern to give greater access to high-quality scientific achievement, Ostwald translated into German the work of the American physicist, Josiah Willard Gibbs, whose papers had only been published in an inaccessible small journal, the *Transactions of the Connecticut Academy of Sciences*, unknown in Europe.

Handbook (Handbuch der allgemeinen Chemie)

Ostwald's encyclopedic *Handbuch der allgemeinen Chemie* was intended to have an international character. Ostwald solicited participation by Frederick Donnan and William Ramsay (British), Arthur A. Noyes (American), Svante Arrhenius (Swedish), and Philippe A. Guye (French). Before the war only volume 2 by Ramsay and G. Rudorf about the noble gases was published. After the war some further volumes were published but without the planned international participation. Ostwald's book *Die chemische Literatur und die Organisation der Wissenschaft* (1919), the first volume of the *Handbuch*, was set in type in 1914 but not printed until 1919. It summarized Ostwald's ideas on the organization of scientific publication and communication (Satoh, 1987). It was probably the first book on chemical literature, although it was not really a literature guide (Mellon, 1982 p. 245).

The "Brücke" (Bridge)—The World Brain

Background: Ostwald's Organizational and International Experience

In 1908 Ostwald wrote, "Everyone who is active in science in any way appreciates the fact that the task of comprehensively organizing scientific reporting or abstracting is a necessity which constantly grows more urgent. Now more than ever this need presents an international aspect and requires the cooperation of various countries" (Ostwald, 1955, p. 374; Ostwald, 1910a, p. 591). Ostwald also pointed out the problem of language and proposed the use of a synthetic auxiliary language as a medium for international communication. (He had been engaged in artificial languages since the beginning of the century.)

Crawford (1992) described the time from 1900 to 1914 as the golden age of internationalism. Ostwald personally participated in many international ventures, especially efforts to set up international networks of various kinds: He was born at the edge of Europe in Latvia; in a sense he was himself an international immigrant to the intellectual heartland. He organized the international development of modern physical chemistry. In 1911 he took part in the foundation of the International Association of Chemical Societies. In 1905 he was the first German exchange professor with the United States. With Emil Fischer and Walther Nernst, he tried to call into being a German Imperial Chemical Institute, the Chemische Reichsanstalt (Johnson, 1990).

Foundation of the Bridge

Ostwald's efforts in scientific publication and his international efforts led to the foundation in 1911 of Die Brücke, Internationales Institut zur Organisierung der geistigen Arbeit (The Bridge, International Institute for the Organization of Intellectual Work) by Wilhelm Ostwald, Karl Bührer, and Adolf Saager (Hapke, 1997).

Karl Wilhelm Bührer and Adolf Saager (1911) published the book *Die Organisierung der geistigen Arbeit durch die Brücke* (The organization of intellectual work through the Bridge). Ostwald's gift of his Nobel prize money made possible the formal opening of the institution, the Bridge, on 11 June 1911.

Because of his international contacts many intellectuals from abroad became members of the Bridge, including the Swedish chemist, Svante Arrhenius; the American industrialist, Andrew Carnegie; the English physicist, Ernest Rutherford; the Swedish writer, Selma Lagerloef; the French mathematician, Henri Poincaré; the Austrian Nobel laureates for peace, Bertha von Suttner and Alfred H. Fried; the Belgian industrialist, Ernest Solvay; the American zoologist and bibliographer, Herbert Haviland Field; and Paul Otlet, a founder of the Institut de Bibliographie in Brussels in 1895.

The term *Gehirn der Welt* (world brain), which Ostwald (1912) claimed the new organization would create, was probably taken from a little book by Alfred H. Fried (1908), which gives a contemporary view of internationalism. Ostwald referred to this book in his periodical *Annalen der Naturphilosophie* (1910b, 9, 194–195), when reviewing the popular serial *Aus Natur und Geisteswelt* published by Teubner. In Fried's little book mention was made of a 1907 article by Friedrich Naumann, "Das Gehirn der Menschheit" (Fried, 1908, p. 28). Like several other members of the Bridge, Naumann was also a member of the German Werkbund.

Prehistory of the Bridge: The "Internationale Monogesellschaft"

Of the Bridge's two other founders, more is known about Karl Wilhelm Bührer than about Adolf Saager. Bührer was born on 1 June 1861, in Bibern (Kanton Schaffhausen, Switzerland) and probably died during or shortly after World War I. He was an editor in Switzerland and moved to Munich in 1908. He founded a so-called Internationale Monogesellschaft in Winterthur as a stock corporation on 27 November 1905 (clipping from an unknown newspaper in Stadtbibliothek Winterthur, Switzerland). The aim of this enterprise was to raise the artistic level of contemporary advertising. One way to accomplish this was the publication of "Monos," little cards or leaflets in a standardized format (Bührer, 1906, backcover). Monos were something like the many *Reklamebilder* (advertising picture-cards) then in circulation in Germany, e.g., from the companies of Stollwerk or Liebig (Selig, 1997).

The "Mono-System" was planned so that the individual monos would complement each other and, collectively, form a well-designed, comprehensive encyclopedia. "The picture side usually contained advertising. The reverse contained a brief statement ('monograph'—that is the reason for the term Mono) explaining the content of the picture, with carefully written advertising slogans of the firms involved in the system" (Das Mono, 1944, p. 253). A box of Monos has survived at the Stadtbibliothek in Winterthur.

According to Ostwald (1926–27, vol. 3, p. 289) Saager was only Bührer's "friend by chance" ("zufälliger Bekannter"). Born in 1879, he studied science and completed a doctoral dissertation in chemistry at Heidelberg in 1902. Later he was active as a writer in Ansbach and in Munich. He wrote a short popular book on chemistry, a city guide about Ansbach, and biographies of Henry Ford, Graf von Zeppelin, and Benito Mussolini. It is probable that he was responsible for the connection with the publisher of most of the Bridge papers, the Seybold'sche Sortimentsbuchhandlung (Seyerlein, 1991). On 31 August 1949 he died in Lugano.

Aims

"Die Brücke is planned as a central station, where any question which may be raised with respect to any field of intellectual work whatever finds either direct answer or else indirect, in the sense that the inquirer is advised as to the place where he can obtain sufficient information" (Ostwald, 1913, p. 6, English original).

The Bridge was supposed to be the information office for the information offices, a "bridge" between the "islands" where all other institutions—associations, societies, libraries, museums, companies, and individuals—"were working for culture and civilization" (Die Brücke, 1910–1911). The organization of intellectual work was intended to occur "automatically" through the general introduction of standardized means of communication—the monographic principle, standardized formats, and uniform indexing (*Registraturvermerke*) for all publications. The following facilities were planned: a collection of addresses, a *Brückenarchiv* as a "comprehensive, illustrated world encyclopedia on sheets of standardized formats," which should contain a world dictionary and a world museum catalog; a *Brückenmuseum*; and a head office and *Hochschule* (college) for organization. "Close cooperation" with the Institut Internationale de Bibliographie in Brussels was also planned.

"Within the last few years successful efforts have been made in America to introduce the idea of scientific management in all sorts of fields, so that we may expect with confidence to find there a responsive audience when we speak of the organization and systematization of the world's intellectual work" (Ostwald, 1913, p. 6). Here Ostwald referred to the work of Frederick Winslow Taylor. According to Burchardt (1977), Ostwald's philosophy influenced the reception of taylorism in Germany, visible in the citations of Ostwald's work in the German edition of *The Principles of Scientific Management* (Taylor, 1919).

Standardization of Paper Formats and the Monographic Principle

Ostwald proposed new standardized formats for all publications. Among the promised advantages of standardizing paper sizes were saving space in desks, bookcases, and libraries; the resultant standardization of printing machines; reduction in the price of publications; as well as the increased feasibility of assembling personal compilations of published materials. One of the Bridge's booklets, *Raumnot und Weltformat* (Bührer & Saager, 1912b), described how a large number of volumes could be shelved in relatively small rooms if their formats were standardized.

In his book about chemical literature (Ostwald, 1919), Ostwald summarized many of the aims of the Bridge and predicted new publication formats. The periodical will be split into separate papers because no scientist wants to read the whole periodical. Ostwald's "Prinzip der unabhängigen Handhabung des einzelnen Stückes" (Principle of the independent use of the individual piece) (Ostwald, 1919, p. 96) was already applied by Ostwald in the publication of his *Klassiker der exakten Wissenschaften*. The principle spoke to "the need to split up scientific communications into very small component parts, which could then at an appropiate instant be built up in any combination and in accordance with the changes that occur with the passage of time in a given area of knowledge" (Bonitz, 1980, p. 29). Paul Otlet had developed similar ideas in 1903 for which he later, in 1918, used the term *monographic principle*.

Ostwald's utopian handbook of the future was intended to be "completely up-to-date at all times" (Ostwald, 1919, p. 93). It is a predecessor of today's looseleaf collections, which in the future will probably be implemented through electronic publishing. Ostwald also applied the principles of the Bridge to his special subject, chemistry, by proposing the foundation of an International Institute for Chemistry (Ostwald, 1914), planned more or less as a "small Bridge."

Influence of Advertising

According to Rayward, "It is possible that Otlet's use of the term [monographic principle] derives from his involvement in Die Brücke" (1994, p. 238). Since the Monos connected with the origins of the Bridge and the Monos were advertising pieces, then one of the important principles of Otlet's contribution to information science may well have originated, at least terminologically, in advertising.

Another interesting connection of the Bridge to advertising was the participation of many members in the Deutsche Werkbund, including Georg Kerschensteiner, Peter Behrens, and Hermann Muthesius (Campbell, 1981, pp. 172–173). One of the most important aims of the Werkbund was to connect art, arts and crafts, and industrial design, which in turn would have some influence on advertising. For the connection of Wilhelm Ostwald to the Werkbund after World War I, see Schirren (1998).

Further Activities

The Bridge published more than twenty leaflets about its aims and activities, and in 1913 the Bridge began its own periodical, the *Brückenzeitung*, edited by Wilhelm Ostwald, Wilhelm Exner, and Karl Wilhelm Bührer. By July 1912, the Bridge had 361 members (BBAW Ostwald papers, No. 3470, Letter of Bührer to Ostwald of 7 July 1912), and the first annual meeting took place in Munich, 28–29 March 1913 (Erste Jahresversammlung, 1913; Première Assemblée, 1913). The Bridge was also involved in exhibitions, for example, in the Bayerische Gewerbeschau in Munich in 1912. It published a list of the world's largest libraries and a translation into German of an excerpt of the decimal classification tables with an index (Bührer & Saager, 1912a).

By 1914 about DM 100,000 or two-thirds of Ostwald's Nobel Prize money was spent. Lack of other funding and organizational problems with Bührer (Ostwald, 1926–27, Vol. 3, pp. 303–306) forced the Bridge to close in 1914. After World War I, Ostwald received a letter from Frank Richard Behrens (Letter of Behrens to Ostwald of BBAW Ostwald papers No. 3470, 1 May 1920), representing an organization called the Bridge in Berlin. It seems there was an attempt to reorganize the Bridge, and Ostwald was asked to become an honorary member.

Connections to Other Bibliographic Activities

Institut International de Bibliographie

In the area of connections and reciprocal influences between Bührer, Ostwald, Otlet, and the Institut International de Bibliographie (IIB), much research remains to be done. (Further research on the connections between these activities and those of Paul Otlet and the Institut International de Bibliographie in Brussels will be possible when the Otlet Papers in the Mundaneum in Mons become accessible.) According to Schneiders (1982, p. 89), the first contact between the Internationale Monogesellschaft and Otlet was in October 1908. Otlet responded enthusiastically to the aims of the Monogesellschaft. They went together well with his universal classification. Using decimal notation on the Mono cards seemed a good way to popularize the decimal classification.

Bührer and Saager (1911) mentioned in the introduction (p. viii) of their programmatic book that there existed an arrangement from 1 May 1911, between the IIB and the Bridge concerning 1) the suitable division of labor between the two (the more scientific part for the IIB, the more practical for the Bridge); 2) the *Weltformate*, a definite scientific scale of size for books and publications; 3) the *äussere Form der Registraturvermerke*, a note on the back cover or inside every book that describes the book, similar to the cataloging-in-publications data now seen in U.S. publications today; and 4) the *Ehrenpräsident* (honorary president) of the Bridge, who would be the *Generalsekretär* (secretary general) of the IIB.

Some differences must have arisen between Otlet and the Bridge. Bührer wrote in a letter to Ostwald on 8 October 1912, that "Hr. Chavannes aus Lausanne," who wanted to found a branch of the Bridge in Switzerland, "formed an alliance with Mr. Otlet." "It would obviously be preferable to me if you take hold of the scepter, because through this a tighter rein can be kept on Mr. Otlet" (BBAW Ostwald papers No. 3470).

The first direct contact between Otlet and Ostwald was probably at the World Congress of International Associations in May 1910. Otlet was one of the Secretaries General of the Congress, while Ostwald and Ernest Solvay were co-chairmen of a section on standardization (Rayward, 1975, p. 180). The personal connection between Ostwald and Otlet may well have been slight, as evidenced by Ostwald's brief mention in his autobiography of Otlet merely as a member of the Bridge (Ostwald, 1926–27, Vol. 3, p. 299).

A postcard from Fried to Ostwald (BBAW Ostwald papers, No. 828, 1 December 1911) points to other interrelations between the international movements before World War I. "I want to call your attention to the new novel *Der Menschheit Hochgedanken* by Baroness Suttner describing a congress of man's outstanding thinkers, which an American multimillionaire decided to hold every year in Luzern. This is an idea that you have already dealt with as Lafontaine just reported to me a few days ago." This novel by Bertha von Suttner contains biographies of the participants at the conference, including a biographical sketch of a man with many of Ostwald's characteristics (Suttner, 1911, p. 166).

Institut für Techno-Bibliographie

Hermann Beck's Institut für Techno-Bibliographie, founded in 1908 (Behrends, 1995, pp. 19–28), is another example of a German organization participating in the bibliographic movements at the beginning of this century. The institute attempted to organize and summarize all forms of technical literature. The names of Beck and Ostwald were also written below an *Aufruf zur Gründung eines deutschen Archivs der Weltliteratur* (Appeal for the establishment of a German archive of the world's literature, 1912), which is reprinted in facsimile in Behrends's book about the history of documentation in Germany until the end of World War II (1995, pp. 231–234).

Hermann Beck was born on 25 August 1879 in Mülheim an der Ruhr. He studied mechanical engineering and social sciences in Dresden, Berlin, and Heidelberg. His publications show his close relationship to social democracy. After World War I, Beck was active in trying to organize the further development of his *Deutsche Archiv der Weltliteratur* using cards for abstracts (Beck, 1919).

In 1905 Beck had already established the Internationales Institut für Sozial-Bibliographie. This institute published the *Bibliographie der Sozialwissenschaften* and the periodical *Kritische Blätter für die gesamten Sozialwissenschaften*, edited by Beck. In this periodical there are several papers by Beck (e.g., Beck, 1907) and others (e.g., Hanauer, 1908) about the contemporary bibliographic movement. The Institut für Techno-Bibliographie was organized in the same way as the Institut für Sozial-Bibliographie. Both intended to combine a subject-oriented central library, a bibliographic card index, an information agency, a bureau of translation, a clipping service, and a bookseller with international coverage (Beck, 1909, p. 113). Beck was also the editor of the periodicals *Technik und Wirtschaft* (Technology and Economy) (1.1908–37.1944, 11-9) and *Dokumente des Fortschritts* (Documents of Progress) (1.1907/08–11.1918, 3), both of which carried some bibliographic items as well.

In a letter to Ostwald, Beck called himself Ostwald's disciple and follower (*"Schüler und Jünger"*) (BBAW Ostwald papers, No. 149, 5 January 1910). In a 27 November 1911 letter Beck enclosed a "Memorial on the Bridge" in which he proposed the union and cooperation of the two enterprises, his Deutsches Archiv der Weltliteratur and Ostwald's Bridge *(Denkschrift betr. ein Zusammengehen des "Deutschen Archivs der Weltliteratur" und des Bibliographischen Zentral-Verlags G.m.b.H., beide in Berlin, mit der "Brücke" in München)* (Beck, 1911). In his memorial Beck also criticized the statutes and the aims of the Bridge. He questioned its requiring by its statutes a very far-reaching connection with the Dewey System and the IIB. For the reception of the IIB and UDC (Universal Decimal System) in Germany see Naetebus (1909), Eichler (1896), and Hanauer (1908 and 1928). Hanauer and Naetebus were the only German participants at the Conférence de Bibliographie in Brussels in 1908. Another important enterprise in bibliography in Germany around the turn of the century was participation in the International Catalog of Scientific Literature (Brodmann, 1901; Tautz, 1903). In the end the cross-purposes of the Bridge and Beck's organizations may be the reason that Beck's plans for cooperation never became reality. On 6 July 1912 Bührer reported to Ostwald that "Beck is supposed to have caused a lot of dubious situations" (BBAW Ostwald papers, No. 3470).

Connection of Wilhelm Ostwald to Julius Hanauer

The correspondence of Ostwald (BBAW) contains more than thirty letters between him and Julius Hanauer, from the year 1895 until 1932, the year of Ostwald's death. The activities of both men in the areas of organization, standardization, and classification are mirrored in these letters.

Julius Hanauer, born 21 September 1872, in Mannheim, studied physics, mathematics and chemistry. After 1896 he worked for four years in industry. He acted as a co-founder of Hermann Beck's Internationales Institut für Sozial-Bibliographie. Between 1908 and 1910 he worked with Otlet at the Institut International de Bibliographie in Brussels. After World War I he was librarian at the Literarisches Bureau of the Allgemeine Elektrizitäts-Gesellschaft in Berlin. He was called "the Saint Francis of the UDC" in Germany (Björkbom, 1978, p. 104). For the role of Hanauer in promoting the development of the UDC in Germany, see also Wimmer (1985). In 1935 he was retired and living in Frankfurt. He died during World War II.

Only seven letters from Hanauer to Ostwald, of the thirty-six kept in the BBAW (Ostwald papers, No. 1072), date from before World War I. In a letter from Brussels (3 March 1912) Hanauer asked Ostwald for printed papers of the Bridge, and he wrote, "As far as I am in a position to judge on this matter, I do not believe, that intellectual work can be organized against the intentions of librarians." Although working for some years in Brussels, Hanauer expressed in the same letter a reservation about Otlet: "I want to be present at the harvest after years of sowing. However I must reject working together with Mr. Otlet."

The meeting of minds between Ostwald and Hanauer is also demonstrated in Hanauer's letter to Ostwald around spring 1920 (the date is unclear). Hanauer wrote about Ostwald's book *Das grosse Elixier* (1920), "My secretary, to whom I had given your book *The Great Elixir* to read, said: 'This is exactly the same as what you say.' "

Conclusion: Wilhelm Ostwald as a Predecessor of Information Science

Being aware of the information problem at the beginning of the twentieth century and looking for alternatives to the scientific journal or for improved means of scientific communication in general, Ostwald and his fellow activists opened a discussion that now at the end of this century continues in the day of the Internet and the proliferation of electronic journals.

The activities of Ostwald and the Bridge concerning the organization of scientific publication and communication had little influence on the scientific community or on the librarians' community in the early decades of this century or after. Ostwald, after his retirement in 1906 and after his support for energetics and monism, was an outsider to the scientific community, even though he had received the Nobel Prize in 1909. Similarly, Ostwald had no close contact with the librarian scene, with the exception of Hanauer, who was himself an outsider because of his support of the decimal classification.

Nevertheless, Ostwald can be seen as a predecessor of information science. Ostwald predicted the arrival of the information specialist as a consequence of the growing division of scientific work. "Therefore, it is ever more necessary for the news service in science, which has been organized up to now in periodicals, annual reports, and similar literary aids, to be built up in such a way that it will be managed by co-workers who are more skillful because specially trained" (Ostwald, 1909, p. 175).

This paper tries to give a picture of the interrelations of a part of the international bibliographic movement before and after World War I. The many similarities between such men as Ostwald, Beck, and Otlet testifies to the existence of a "bibliographic movement" at the beginning of this century. This movement was noticed in Germany, but it only very slowly changed the thinking about the importance of technical and scientific literature in the minds of German librarians, who were mainly trained in the humanities.

References

Beck, H. (1907). Die internationale Bibliographie und ihre Zukunft [The international bibliography and its future]. *Kritische Blätter der gesamten Sozialwissenschaften*, *3,* 203–212.

Beck, H. (1909). Zur Organisation der technischen Auskunft: Dem neugegründeten Internationalen Institut für Techno-Bibliographie auf den Weg [Organization of technical information: Preparing the way for the newly established International Institute for techno-bibliography]. *Technik und Wirtschaft*, *2,* 110–117.

Beck, H. (1911). Denkschrift betr. ein Zusammengehen des "Deutschen Archivs der Weltliteratur" und des "Bibliographischen Zentral-Verlags G.m.b.H.," beide in Berlin, mit der "Brücke" in München [Memorandum concerning merging the "German archive of world literature" and the "Bibilographic Central Publishing Office," both in Berlin, with the "Bridge" in Munich]. BBAW Ostwald papers No. 149.

Beck, H. (1919). *Zur Organisierung des Wissenschaftsbetriebes, insbesondere des Schriftwesens* [Organization of scholarly work, especially publication issues]. Published with W. Borgius, *Zur Sozialisierung* *des Buchwesens* [Nationalization of publication issues] (pp. 42–53). Berlin: Verlag Neues Vaterland.

Behrends, E. (1995). *Technisch-wissenschaftliche Dokumentation in Deutschland von 1900 bis 1945 : unter besonderer Berücksichtigung des Verhältnisses von Bibliothek und Dokumentation* [Technical and scientific documentation in Germany from 1900 to 1945: with special reference to the relationship between library and documentation]. Wiesbaden: Harrassowitz.

Berlin-Brandenburgische Akademie der Wissenschaften (BBAW), Akademie-Archiv, Nachlass Ostwald [Ostwald papers].

Björkbom, C. (1978). "The History of the Word Documentation within the FID." In P. R. Frank. (Ed.), (1978). *Von der systematischen Bibliographie zur Dokumentation* [From systematic bibliography to documentation] (pp. 103–108). Darmstadt : Wissenschaftliche Buchgesellschaft.(Original work published 1959).

Bonitz, M. (1979). *Wissenschaftliche Forschung und wissenschaftliche Information* [Scientific research and scientific information]. Berlin: Akademie-Verlag.

Bonitz, M. (1980). Wilhelm Ostwald on scientific-information problems. *Scientific and Technical Information Processing, 7*, 28–32.

Brodmann. (1901). Der internationale Katalog der naturwissenschaftlichen Literatur [The international catalog of scientific literature]. *Zentralblatt für Bibliothekswesen*, *18*, 493–500.

Die Brücke (1910–1911). *Annuaire de la Vie Internationale*, 1910–1911, pp.623–632 (with the Statutes of the Bridge).

Buckland, M., & Plaunt, C. (1997). Selecting libraries, selecting documents, selecting data. In *Proceedings of the International Symposium on Research, Development & Practice in Digital Libraries 1997*. Available: http://bliss.berkeley.edu:80/papers/isdl97/isdl97.html

Bührer, K. W. (1906). *Schweizerbier* [Swiss beer]. Winterthur: Verlag der Internationalen Monogesellschaft.

Bührer, K. W., & Saager, A. (1911). *Die Organisierung der geistigen Arbeit durch "Die Brücke."* [The organization of intellectual work through the "Bridge"]. Ansbach: Fr. Seybold's Buchhandlung.

Bührer, K. W., & Saager, A. (1912a). *Die Welt-Registratur: Das Melvil-Deweysche Dezimalsystem* [The world registry: The Melvil Dewey Decimal system]. Munich: Die Brücke.

Bührer, K. W., & Saager, A. (1912b). *Raumnot und Weltformat* [Lack of space and the world format]. Munich: Die Brücke.

Bugge, G. (1925). Die Auswertung der technischen Literatur [Indexing technical literature]. *Zeitschrift des Vereins Deutscher Ingenieure*, *69,* 1517–1522.

Burchardt, L. (1977). Technischer Fortschritt und Wandel: Das Beispiel der Taylorismus-Rezeption [Technical progress and change: the case of the reception of Taylorism]. In W. Treue (Ed.), *Deutsche Technikgeschichte* [The history of German technology] (p. 52ff). Göttingen: Vandenhöck & Ruprecht.

*Campbell, J. (1981). *Der Deutsche Werkbund: 1907–1934* [The German Werkbund]. Stuttgart: Klett-Cotta.

Crawford, E. (1992). *Nationalism and internationalism in science, 1880–1939*. Cambridge: University Press.

Dunsch, L. (1989). *Ein Fundament zum Gebäude der Wissenschaft: Einhundert Jahre Ostwalds Klassiker der exakten Wissenschaften (1889–1989)* [The base for the edifice of science: One hundred years of Ostwald's classics of exact sciences (1889–1989)]. Leipzig: Akadem. Verlagsgesellschaft. Geest & Portig.

Eichler, F. (1896). Der Kampf um das Dewey'sche System [The controversy about Dewey's system]. *Zentralblatt für Bibliothekswesen*, *13,* 424–425.

Erste Jahresversammlung der Brücke: am 28. und 29. März 1913 [First annual meeting of the"Bridge": 28 and 29 March 1913]. *Brückenzeitung*, No. 8 and 9, 1–18.

Fleck, G. O. (1993). Wilhelm Ostwald, 1853–1932. In L. K. James (Ed.), *Nobel laureates in chemistry 1901–1992* (pp. 61–66). Washington, DC: American Chemical Society.

Fried, A. H. (1908). *Das internationale Leben der Gegenwart* [International life today]. Leipzig: Teubner.

Hakfoort, C. (1992). Science deified: Wilhelm Ostwald's energeticist world-view and the history of scientism. *Annals of Science*, *49,* 525–544.

Hanauer, J. (1908). Eine internationale bibliographische Konferenz [An international bibliographic conference]. *Kritische Blätter der gesamten Sozialwissenschaften*, *4,* 99–102.

Hanauer, J. (1928). Dezimalklassifikation [Decimal classification]. *Zentralblatt für Bibliothekswesen*, *45,* 503–514.

Hapke, T. (1997). Wilhelm Ostwald und seine Initiative zur Organisation und Standardisierung naturwissenschaftlicher Publizistik: Enzyklopaedismus, Internationalismus und Taylorismus am Beginn des 20. Jahrhunderts [Wilhelm Ostwald and his initiatives for the organization and standardization of scientific journals: Encyclopedism, internationalism and Taylorism at the beginning of the 20th century]. In C. Meinel (Ed.), *Fachschrifttum, Bibliothek und Naturwissenschaft im 19. und 20. Jahrhundert* [Specialized literature, the library and science in the 19th and 20th centuries] (pp. 157–174). Wiesbaden: Harrassowitz.

Hapke, T. (1990). *Die Zeitschrift für physikalische Chemie: Hundert Jahre Wechselwirkung zwischen Fachwissenschaft, Kommunikationsmedium und Gesellschaft* [The *Zeitschrift für physikalische Chemie*: 100 years of interaction between specialized science, communication medium, and society]. Herzberg: Bautz.

Hiebert, E. N., & Körber, H. G. (1978). Ostwald, Friedrich Wilhelm. In C. C. Gillispie, (Ed.), *Dictionary of scientific biography* (Vol. 15, pp. 455–469). New York: Charles Scribner's Sons.

Holt, N. R. (1970). A note on Wilhelm Ostwald's energism. *Isis*, *61,* 386–389.

Holt, N. R. (1977). Wilhelm Ostwald's "The Bridge." *British Journal for the History of Science*, *35,* 146–150.

Homer, T. J. (1912). The Boston Co-operative Information Bureau. *Library Journal*, *37,* 501–504.

Johnson, J. A. (1990). *The Kaiser's chemists: Science and modernization in Imperial Germany*. Chapel Hill: University of North Carolina Press.

Krohn, W., & Küppers, G. (1989). *Die Selbstorganisation der Wissenschaft* [The self-organization of science]. Frankfurt am Main: Suhrkamp.

Leegwater, A. (1986). The development of Wilhelm Ostwald's chemical energetics. *Centaurus*, *29,* 314–337.

Lewandrowski, P. (1979). Der Kampf Wilhelm Ostwalds um die Schaffung eines einheitlichen Informations- und Dokumentationssystems der Wissenschaft—"Die Brücke" [Wilhelm Ostwald's fight for the creation of a uniform information and documentation system for science—"The Bridge"]. In *Internationales Symposium anlässlich des 125. Geburtstages von Wilhelm Ostwald* [International symposium on the occasion of Wilhelm Ostwald's 125th birthday] (pp. 149–156). Berlin: Akademie-Verlag.

Lyons, F. S. L. (1963). *Internationalism in Europe 1815–1914*. Leyden: Sythoff.

Mellon, M. G. (1982). *Chemical publications: Their nature and use* (5th ed.) New York: McGraw-Hill.

Das Mono. (1944). In *Die Lithographie in der Schweiz* [Lithography in Switzerland] (pp. 253–254). Bern: Verein Schweizer Lithographiebesitzer.

Naetebus, G. (1909). Die Brüsseler Conférence Internationale de Bibliographie 1908 [The Brussels International Conference on Bibliography 1908]. *Zentralblatt für Bibliothekswesen*, *26,* 309–313.

Ostwald, W. (1885–1887). *Lehrbuch der allgemeinen Chemie* [Textbook of general chemistry]. 2 vols. Leipzig: Engelmann.

Ostwald, W. (1903a). Wissenschaftliche Massenarbeit [Scholarly mass-scale work]. *Annalen der Naturphilosophie*, *2,* 1–28.

Ostwald, W. (1903b). The relations of biology and the neighboring sciences. *University of California Publications, Psychology*, *1*(4), 11–31.

Ostwald, W. (1909). *Energetische Grundlagen der Kulturwissenschaft* [Energetic foundations of sociology]. Leipzig: Klinkhardt.

*Ostwald, W. (1910a). Berzelius' Jahresbericht und die Organisation der Chemiker (1908) [Berzelius' Jahresbericht and the organization of chemists]. In W. Ostwald, *Die Forderung des Tages* [The demands of the day] (pp. 586–594). Leipzig: Akadem. Verlagsgesellschaft.

Ostwald, W. (1910b). Aus Natur und Geisteswelt [Nature and the intellectual world](Review). *Annalen der Naturphilosophie, 9,* 194–195.

Ostwald, W. (1912). *Das Gehirn der Welt* [The world brain]. Munich: Brücke.

Ostwald, W. (1913). Scientific management for scientists. "The Bridge." The trust idea applied to intellectual production. *Scientific American*, *108,* 5–6.

Ostwald, W. (1914). Memorial on the foundation of an International Chemical Institute. *Science, N.S.*, *40,* 147–156.

Ostwald, W. (1919). *Handbuch der allgemeinen Chemie* [Handbook of general chemistry]: Vol. 1. *Die chemische Literatur und die Organisation der Wissenschaft* [Chemical literature and the organization of science]. Leipzig: Akademische Verlagsgesellschaft.

Ostwald, W. (1920). *Das grosse Elixier: Die Wissenschaftslehre* [The great elixir: The theory of science]. Leipzig-Gaschwitz: Dürr & Weber.

Ostwald, W. (1926-1927). *Lebenslinien* [Life lines]. 3 vols. Berlin: Klasing.

Ostwald, W. (1929) *Die Pyramide der Wissenschaften: Eine Einführung in wissenschaftliches Denken und Arbeiten* [The pyramid of the sciences: An introduction to scientific thinking and working]. Stuttgart: Cotta.

Ostwald, W. (1955). Berzelius' "Jahresbericht" and the international organization of chemists. *Journal of Chemical Education, 32,* 373–375.

Pohle, M. (1998). Wilhelm Ostwald und die Zeitschrift für physikalische Chemie. *Mitteilungen der Wilhelm-Ostwald-Gesellschaft zu Grossbothen*, *3*(3), 40–63.

Première Assemblée annuelle de l'association "Die Brücke" in 1913 [First annual congress of the "Bridge" association in 1913]. (1913). *La vie internationale, 3,* 451–452.

Rayward, W. B. (1975). *The universe of information: The work of Paul Otlet for documentation and international organization*. Moscow: VINITI & FID.

Rayward, W. B. (1994). Visions of Xanadu: Paul Otlet (1868–1944) and hypertext. *Journal of the American Society for Information Science*, *45,* 235–250.

Rayward, W. B. (1997). The origins of information science and the International Institute of Bibliography/International Federation for Information and Documentation (FID). *Journal of the American Society for Information Science*, *48,* 289–300.

Rodnyj, N. I. & Solov'ev, J. I. (1977). *Wilhelm Ostwald*. Leipzig: Teubner.

Satoh, T. (1987). The Bridge movement in Munich and Ostwald's *Treatise on the Organization of Knowledge. Libri*, *37,* 1–24.

Schirren, M. (1998). Vereinheitlichung und Weltformat: Wilhelm Ostwalds Farbnormen und der Traum von der Ordnung der Dinge [Standardization and world format: Wilhelm Ostwald's color standards and the dream of the order of things]. *Archithese*, *28,* 10–13.

*Schneider, G. (1923). *Handbuch der Bibliographie*. Leipzig: Hiersemann.

Schneiders, P. (1982). *De Bibliotheek- en Documentatiebeweging 1880–1914: Bibliografische Odernemingen rond 1900* [The library and documentation movement 1880–1914: Bibliographic enterprises around 1900]. Doctoral dissertation, University of Amsterdam Letterenfaculteit, Amsterdam, Netherlands.

Selig, R. A. (1997 June/July). Reklamebilder—The Germans' Baseball Cards. *German Life*. Available: http://langlab.uta.edu/GERM/German_Life/9706_02.html.

Servos, J. W. (1990). *Physical chemistry from Ostwald to Pauling: The making of a science in America*. Princeton: University Press.

Seyerlein, F. (1991). *Festschrift 275 Jahre: 5. September 1716–5. September 1991: Geschichte der Fr. Seybold´s Sortimentsbuchhandlung in Ansbach* [Volume commemorating 275 years (5 September 1716–5 September 1991): History of Fr. Seybold's retail book trade in Ansbach]. Ansbach: Schmidt.

*Suttner, B. v. (1911). *Der Menschheit Hochgedanken: Roman aus der nächsten Zukunft* [Mankind's higher thoughts: A novel of the near future]. Berlin: Verlag der Friedens-Warte.

Tautz, K. (1903). Das dem Reichsamt des Innern unterstellte Deutsche Bureau der internationalen Bibliographie in Berlin [The German bureau of international bibliography in Berlin, placed under the state department of interior]. *Zentralblatt für Bibliothekswesen*, *20,* 285–290.

Taylor, F. W. (1919). *Die Grundsätze wissenschaftlicher Betriebsführung* [The principles of scientific management]. Munich: Oldenbourg. Reprinted: (1983) 2. ed. Munich: Raben.

Wells, H. G. (1938). *World brain*. London: Methuen.

Wimmer, F. (1985). Die Internationale Dezimalklassifikation: Betrachtungen zu ihrer Entstehungsgeschichte und Aktualität [The universal decimal classification: Observations on the history of its origins and relevance to the present]. *DK-Mitteilungen*, *29*(5), 15–18.

***References Available in English Translation**

Campbell, J. (1978). The German Werkbund. Princeton, NJ: Princeton University Press.

Ostwald, W. (1955). Berzelius' "Jahresbericht" and the international organization of chemists. *Journal of Chemical Education*, *32,* 373–375.

Schneider, G. (1934). Theory and history of bibliography, Ralph R. Shaw (Trans.). New York: Columbia University Press.

Suttner, B. v. (1914). *When thoughts will soar: A romance of the immediate future*. Boston: Houghton Mifflin.

Ralph Shaw and the Rapid Selector

Jana Varlejs

Abstract

The Rapid Selector, developed by Vannevar Bush in the 1930s, represented an early attempt to automate document retrieval using photoelectric cells, microfilm, and high-speed photography. It was not until the late 1940s, however, that a librarian attempted to adapt the machine to assist in producing a major bibliographic tool, the *Bibliography of Agriculture*. As director of the library of the U.S. Department of Agriculture and member of the American Documentation Institute, Ralph Shaw understood the need for providing timely access to the burgeoning literature for a widely dispersed scientific and technical community. The Rapid Selector looked like a solution to the problem, but turned out to be a serious disappointment. Shaw's experience with the Selector affected his thinking about automation and led him to warn the profession against a too-ready belief in the promise of machine-assisted retrieval.

Introduction

In the 1930s and early 1940s the mix of scientists and librarians involved in the American Documentation Institute (ADI, the precursor of the American Society for Information Science) shared a concern for making the burgeoning scientific and technical literature accessible. Working together, they advanced microfilming as the technology to solve the dissemination problem and strove to publish timely alerting and indexing services to provide intellectual access to scientific information. For example, in the 1941–42 fiscal year, the Army Medical Library filled "6,208 orders from 1,198 customers requiring exposure of 3 miles of film" (Miles, 1982, p. 300). But World War II and Cold War demands on information handling raised the stakes and accelerated experimentation with new tools and methods. Machine-assisted indexing, storage, retrieval, and dissemination of scientific information became the ultimate goals. Librarians for the most part seldom could afford emerging technology, and opportunities to shape its development were limited. There were of course exceptions, and the subject of this paper is one of those. The technology in this case is the Rapid Selector, the first machine designed specifically for bibliographic retrieval. The librarian is Ralph Shaw, head of the U.S. Department of Agriculture (USDA) library at the time of his involvement with the Selector. The focus here will be more on Shaw than on the machine, and particularly on the impact that his experience with the Selector had on his subsequent thinking about automation. Because Shaw was a leader in the library profession, an educator, a prolific writer, and a frequent consultant and speaker, his opinions were widely known and had considerable influence.

The Rapid Selector

The Selector has been of interest to historians of information science because of its kinship to the Memex, Vannevar Bush's fantasy of a personalized scholar's workstation (Nyce & Kahn, 1991). Bush designed the first version of the Selector in the 1930s, combining photoelectric cells, microfilm, and high-speed photography. While he had a genuine interest in contributing to the solution of the literature control problem, his strongest motivation was to obtain sponsors and funding so that he could support his students and young engineers at MIT. For accounts of the lengthy and complicated history of the Selector, see Burke (1991, 1994), Nyce and Kahn (1991), and Buckland (1992).

The system was designed basically to store documents or abstracts together with coding on microfilm. Searching was done with an interrogating device, such as a punched card or paper tape. When photoelectric cells registered congruence between the inquiry code and the microfilm code, a camera would shoot the appropriate frame and record it onto another film for reproduction and enlargement. Each document could be coded with multiple identifiers. Bush may have imagined a kind of indexing that would realize his dream of a mechanism that would allow association of topics from disparate areas. Both mechanical and conceptual failures dogged the machine throughout its various incarnations. Nevertheless, Bush continued to hope for another opportunity to produce a functional and commercially viable machine.

When Shaw entered the picture, it was in the hope that the Selector might be a more efficient bibliographic tool than his printed *Bibliography of Agriculture* and similar indexes. He was also concerned with the problem of providing access to the "tens of thousands of tons" of scientific and technical U.S. wartime publications that were being declassified, plus material captured from enemies. With expertise in photographic technology and bibliography, commitment to serving the needs of scientists, awareness of the explosion of scientific information in the post–World War II era, and a drive to innovate, Shaw could not resist the idea of a machine that used a combination of microfilm, electronics, and high-speed photography to store, retrieve, and copy bibliographic information. In 1946 he wrote to Vannevar Bush, referring to a 1940 document describing the Rapid Selector, and asked whether he could borrow the prototype. He explained that he wanted "to experiment with its application to the organization of knowledge in a great research library" (Nyce & Kahn, 1991, p. 114). Bush gave his consent, funding was obtained from the Office of Technical Services (OTS) of the Department of Commerce, and the machine was built under the supervision of engineers who had worked on the earlier model at MIT. The new Selector was delivered to the USDA library in 1949.

Shaw's Background

One could forgive Shaw thinking of himself as the right man in the right place at the right time. He developed his interest in scientific information when he worked for the science and technology department of the Cleveland Public Library while attending college at Western Reserve University (biographical information is taken from Stevens, 1978, and Turner, 1983). After obtaining a bachelor's of science in library service at Columbia in 1929, he became chief bibliographer of the Engineering Society Library. He went on for a master's degree at Columbia, writing his thesis on engineering books that were available in America before 1930. In 1934 his translation of Georg Schneider's *Theory and History of Bibliography* from the German was published by Columbia.

During a four-year term as a public library director (1936–40) Shaw began to apply photography to library operations. The result was the Photocharger, a machine for circulation control, although the concept of transaction charging was what Shaw took pride in, rather than the machine that facilitated it (Shaw, 1939; Hines, 1975, p. 9). His interest in photography for management tasks found ample expression when he assumed the directorship of the USDA library in 1940. As recounted by Hines, he streamlined the production of the major index to agricultural information:

> In order that researchers in the field would be helped rapidly to find out what existed so they could request it, Shaw used photography and lithography to produce the *Bibliography of Agriculture*. It was produced by photographing the original typed index cards, laid out shingled on page layout boards. It was a typical Shaw product. It looked like hell, it was done by a tiny staff, but it often left the printer for the subscriber within five days after the last article indexed had been received, and it covered a hundred thousand items a year. The *Bibliography of Agriculture* in those days neatly combined current awareness and retrospective searching values before the term for the first had even been thought of. (1975, p. 7)

Hines goes on to describe Shaw's other uses of photography over the course of his career, ranging from a photostat device that simplified clerical routines before the advent of photocopying, to the use of miniprint to produce publications otherwise too expensive to publish. The photostat machine, called Photoclerk, was developed for use at the USDA library, but Shaw involved eleven other libraries in an experiment to test applications (Shaw, 1953). In reporting on this project, he highlighted not only the savings but also the improvements in management that resulted: "The very existence of an experiment made it necessary to think through policies, programs, and procedures, for . . . this frequently led to broadening of programs or changing of procedures without the use of the camera" (1953, p. 15).

Shaw was by no means the first to apply photography to library operations. The Engineering Society Library, where he worked for seven years, used photostats as early as 1912 (Farkas-Conn, 1990, p. 33). What was creative about Shaw, however, was his ability to take a systems view and to see how a tool could contribute to his ideal of "scientific management." Much taken with Frederick Winslow Taylor, an early-twentieth-century management theoretician, and others who promulgated this approach, he compared it to operations research (Shaw, 1954). He made it a basic principle to scrutinize the purpose of policies and programs and to collect data on the routines and procedures in order to determine their effectiveness and efficiency. One of his famous aphorisms was "do not do efficiently that which does not need to be done" (1958, p. 5).

When he took over the USDA library in 1940, microfilm became another aspect of photographic technique in which he developed expertise. He inherited an arrangement that his predecessor, Claribel Barnett, had made in 1934 (Farkas-Conn, 1990, pp. 41–42) with Watson Davis and others who saw the promise of microfilm in advancing scholarly communication. Barnett's interest grew out of the need to improve upon interlibrary loan and facsimile copies as the primary means to serve the information needs of widely dispersed users at agricultural experiment stations and laboratories. With the introduction of Bibliofilm, as the service came to be called, the library reached beyond its own collection to find, film, and deliver the required document to the user. In the first six months of the project over 150,000 pages were filmed, despite the fact that the service was not promoted and current literature awareness was minimal. The service remained at the USDA until 1941, when Bibliofilm as a part of Science Service and ADI ceased as a centralized operation. While Shaw was director from 1940 to 1954, the USDA library continued a modified relationship with ADI, as well as providing the service for its own clientele (Farkas-Conn, 1990, pp. 88–89). In addition, in 1946, the library cooperated with the American Chemical Society to provide copies of articles in *Chemical Abstracts*, a project that was said to be "of inestimable value in the promotion of research in chemistry" (Mohrhardt, 1957, p. 76). Mohrhardt states that to improve the efficiency of these substantial filming operations, Shaw introduced a camera in 1943 that could be used in the stacks, thus eliminating the need to pull and reshelve materials. This involvement with massive copying probably led him to select copyright as the topic of his dissertation at the University of Chicago, which he completed in 1950.

From 1944 to 1946 Shaw was on leave from the USDA library and served in the Army Air Force Medical Department. Recruited to the Army Medical Library, he worked with Francis St. John to reorganize and streamline operations in time to meet the extraordinary demands for medical literature made upon the library by the military during the war, reaching over 6.5 million pages of microfilm in 1945 (Miles, 1982, pp. 295, 301). Shaw had met Vannevar Bush by 1945 at the latest, when he advised Bush, then chairman of the Office of Scientific Research and Development, to persuade the government to establish an agency that would deal with the mountain of technical and scientific information generated by both the Allies and their enemies (Farkas-Conn, 1990, p. 111). Bush succeeded, and the Publication Board (on which Shaw served) was established under the auspices of the Department of Commerce's Office of Technical Services, headed by John C. Green. It was their mutual interest in the dissemination of "the prodigious store of useful knowledge developed during the last five years under the stress of emergency conditions" (Shaw, 1946, p. 105) that brought Shaw and Green together in pursuit of a machine that would help in the task.

Because Shaw knew ADI's Watson Davis and others who shared the conviction that microfilm was the solution to the storage and dissemination of information, he may have heard of the original Selector well before he found the 1940 document and approached Bush for permission to borrow the prototype. As recounted by Farkas-Conn (1990, p. 19), Davis met Bush in 1932, and the idea for a machine very like the one that became the Bush Selector may have originated in Davis's circle. Burke (1994, p. 43) believes that the basic Selector concept was already in Bush's mind in the early 1930s. In any case, by 1946, the original machine, which had been put on mothballs in 1940 and—according to Bush—had been cannibalized, would have had to be rebuilt if the money could be found (Burke, 1994, p. 334). It was at this point that Shaw's connection with John Green and his Office of Technical Services proved fortuitous, as Green was the key to financing the machine.

Shaw and the Rapid Selector

One suspects that Shaw's curiosity about the Selector, together with the urgency of dealing with unprecedented quantities of information, clouded his usually systematic approach to experimentation. He must have been aware of the specifications of the earlier machine, if not of all the mechanical problems, and should have been able to anticipate the time and cost factors intrinsic to the machine's design. His enthusiasm led to publications describing the Selector before it had been rigorously tested (e.g., Shaw, 1949a; 1949b). By the time Shaw delivered the 1950 Windsor lecture, "Machines and the Bibliographical Problems of the Twentieth Century," he had begun to think not only about the cost-benefit aspects but also about the need for a systems approach: "Until we know what we are trying to achieve, how, why, and for whom, and the amount of effort which may justifiably be assigned to the solution of these problems, it will not be possible to design machines to solve the mechanical problems, nor will it be possible to use existent machines intelligently" (Shaw, 1951a, p. 70).

Meanwhile, building the Selector had turned into a cliff-hanger. The economic, engineering, political, and patent problems are described in detail in Burke's *Information and Secrecy* (1994, chap. 13). It took personal intervention from Vannevar Bush to prod his protégés at the engineering firm that held the contract to complete the project. Burke (1994, p. 345) suspects that the engineers may have realized that the Selector design was already obsolete and stalled in order not to embarrass Bush. Ironically, Bush prevailed, and the machine was delivered to Shaw at the USDA library in 1949, where it failed to work. To add to the dismay, the patent office discovered the claim of Emanuel Goldberg, who had patented a design very similar to that of the Bush machine in 1931 (on Goldberg, see Buckland, 1992). Shaw also became aware of the claim when Goldberg, having learned of the Rapid Selector's debut at the USDA library, paid a visit (Buckland, 1992, p. 58). Shaw gave recognition to Goldberg in some of his writing after that, notably in the Windsor lecture (Shaw, 1951a, p. 58). Bush, however, never acknowledged Goldberg, although it is known that he had been informed about him (Zachary, 1997, p. 265).

Between 1949 and 1952, when Shaw gave up on the Selector, work on the machine continued. Shaw, Bush, and the National Bureau of Standards engineers made modifications and rebuilt parts of the machine in an attempt to save it, but these efforts did not make the Selector functional for Shaw's purposes. As reported by Bagg and Stevens:

> The major factor causing abandonment of the machine was that it was not designed to copy successive frames without delays that severely increased search time. Moreover, the limitation of the selection code area to six selection criteria per document frame and the limitation of the question to one criterion per run had seriously restrictive effects upon indexing and search, and therefore upon the practical use of the selector. (1961, p. 23)

In the opinion of another critic, Scott Adams, the Selector could not be effective because "Shaw had not grappled with the fundamental problems of indexing, so critical for information retrieval" (Farkas-Conn, 1990, p. 134). Adams, Shaw's colleague as one of the librarians recruited to serve during wartime at the Army Medical Library, was certainly qualified to make this judgment. His concern about the inconsistency of subject headings in the various publications providing bibliographic control of the medical literature led him to organize a conference on the problem in 1947 (Miles, 1982, p. 390). It was Shaw rather than Adams, however, who was appointed in the following year by Raymond Bliss, surgeon general of the Army, to serve on a Committee of Consultants for the Study of the Indexes to the Medical Literature Published by the Army Medical Library. Thanks to a research group attached to this committee, important progress was made in using punched cards to produce a subject heading authority list (Miles, 1982, p. 339). By the time the committee finished its work in 1950, the Rapid Selector may have been beyond the point where Shaw could have applied the research results to the machine's redesign. He might not have wanted to tinker with the indexing in any case, since he seemed to have a blind spot when it came to knowledge representation. Despite his association with many of those who were deeply involved in thinking about and developing indexing and coding schemes during this era, Shaw did not appear to have a solid grasp of the subject. Frederick Kilgour (personal communication, September 1998) and Winifred Sewell (personal communication, October 1998) confirm Adams's opinion of Shaw's failings in this regard. Sewell, who worked on revising medical subject headings to be used in the first computerization of *Index Medicus*, recalls that Shaw failed to understand the details of how MEDLARS (Medical Literature Analysis and Retrieval System) worked. Thus, it is understandable that Shaw's publications about the Selector focused on the mechanical problems and the length of time that it took to perform a search, while avoiding any in-depth discussion of the indexing and coding difficulties.

Perhaps for the first time, Shaw was faced with a major failure. What may have been especially galling was the realization that the problem with the Selector was not simply one of inadequate engineering or mechanics. Rather the neglect of what should have been the first step—a rigorous examination of indexing and searching in the machine context—was at least as much at fault. As there seems to be no contemporary record of

Shaw's thinking in regard to the indexing and coding scheme for the Selector at the time that he developed it, one can only speculate on the basis of what he wrote later. He stated that "a really important contribution to the advancement of science will result only if we can rethink the methods of organization of knowledge to take full advantage of the new technique . . . We need first to do some fundamental thinking and some operational research to determine what is really needed for the advancement of scientific communication" (Shaw, 1951a, p. 66). He most likely was thinking in terms of studying users rather than tackling subject access.

He goes on to talk about the feasibility of using uncontrolled vocabulary in the machine context, allowing for the development of new discoveries, as there is not the same limitation to the number of descriptors as in manual systems. Here he seems to be kowtowing to Vannevar Bush, who disliked the hierarchical, controlled systems used by librarians (Nyce & Kahn, 1991, pp. 117–118; Burke, 1994, p. 190). Writing elsewhere, Shaw saw Bush's vision of indexing by association essentially as fantasy:

> [Machines] do not now offer any promise whatsoever for elimination of the intellectual effort involved in bibliographic work; and fuzzy thinking about the creation of new knowledge by assembling unrelated data mechanically is probably responsible for a large part of the delay in applying machine techniques to the parts of the job they may be able to handle. Tools and machines of some types appear to be indispensable and have always been used for storage, selection, and reproduction of bibliographic materials. Those aspects of the problem appear to constitute the field of application of machines. Machines do not now, nor will they in the foreseeable future, handle the intellectual aspects of bibliography. (1951b, pp. 201–202)

While Shaw recognized the intellectual challenge of indexing, he was too much of a pragmatist and too grounded in his own experience as a librarian to be able to jettison traditional principles of classification and subject access in favor of new approaches. He was used to the model of the *Bibliography of Agriculture*, which allowed one to browse broad categories or to zero in on very specific subjects (Olivieri & Forbes, 1969, p. 451). The early volumes of the *Bibliography* illustrate the dependence on classification to offset rather rudimentary and somewhat careless indexing. Shaw emphasized speed in preparing and distributing the publication to the detriment of the quality of subject access. In a 1956 speech Shaw referred to the conflict between a desire to draw together concepts from disparate fields and the ability to scan categories within a field. He stated that in designing the indexing and coding for the Selector:

> The basic error was the assumption that we could run fast enough to avoid pre-classification; yet in terms of the total amount of material in a research library, this experiment showed the futility of running instead of thinking. There appears to be no reason for running all ancient history when we are looking for something in gamma-ray physics and an order of at least 1,000 times the net speed can be achieved merely by the roughest sort of pre-classification by broad subjects and periods. This would make it possible to use 50-ft cartridges instead of 2,000-ft rolls, and to change the search time from six-minute units to half-minute or one-minute units. This requires additional development work, but the principle has been established. (1958, p. 31)

Here he seems to be saying that it is unrealistic to expect the machine to permit efficient searching of a very large database containing unrelated subjects. He does not, however, clearly state the other problem with the particular version of the Selector that he had tested, which was that his indexing and coding scheme, together with the way the machine was constructed, required an exact match between an inquiry and the item indexed (Burke, 1994, pp. 189, 340; Jahoda, 1961, pp. 175–176). Because the "selected abstracts could not be re-run through the Rapid Selector . . . it could not be used for conducting a search whose scope might require more than one characteristic for definition" (Perry, Kent, & Berry, 1956, p. 53). Carl Wise and James Perry had made suggestions for improving the coding, while Calvin Mooers proposed his own Zatocoding (Jahoda, 1961, pp. 177–178). Shaw seems not to have reacted to these proposals, while Bush did not concede the critical nature of coding until the 1960s (Zachary, 1997, pp. 272–273).

In addition to his blindness in regard to indexing, another reason for Shaw's failures with the Selector was his departure from his own habit of looking at the total system, analyzing it in terms of purpose and effectiveness, discarding what was superfluous, and finding or creating the tool to do the job efficiently. The transaction system that he invented while at the Gary Public Library, the USDA library's Photoclerk, and the production method for the *Bibliography of Agriculture* arose from his identification of specific problems in particular systems that called for economical solutions. The Rapid Selector does not fit this pattern. It was someone else's solution to a problem, and it is doubtful that Shaw would have placed faith in it had it not been backed by the highly respected Bush, who originally conceived it at MIT as a successor to an analog calculator for purposes of data retrieval. Exactly how the basic idea would be realized depended on who funded the machine (Burke, 1991). While the vision of the Memex probably hovered in the background, Bush never systematically studied how to build search-and-retrieval logic into the machine. He missed the opportunity to give it "and/or" searching capability, gave short shrift to problems of coding and indexing, and gave priority to making the machine run at the greatest possible speed (Burke, 1994, pp. 189–191), a priority that resonated with Shaw.

Once Shaw had hands-on contact with the machine, he concentrated on the mechanical rather than the intellectual problems. He produced two patents, one related to eliminating double exposures when two hits were too close together and the other to the camera used to create microfilm from varying-sized text together with standard-sized codes (Jahoda, 1961, p. 183; Shaw, 1950). While he continued to advocate use of the Rapid Selector for several years after its initial failure, he qualified his support by pointing out that in order for it to become useful considerable research was needed on how to organize information for machine sorting. He emphasized the need to consider the entire system time and cost (coding, preparation of the interrogating mask, developing the search results film) as opposed to allowing speed of sorting to tempt one into thinking of the machine as efficient. Having been beguiled himself by specifications for a machine that used what appeared to be familiar photographic technology and added the attraction of high speed, he could issue the warning with conviction.

The Aftermath

By 1953 Shaw was reminding librarians that the book was still the most efficient tool for storing and finding information; that machine solutions were proposed too glibly for solving exaggerated problems; and that it would take librarians, not outsiders, to develop a better bibliographic tool, electronic or not:

> So developing new tools will always be a part of our jobs. If they are to be electronic, well and good. If not, well and good. But each will have to justify itself by more than catchwords and will have to serve as more than a development project. If they do not, they are gadgets rather than tools. (1976, p. 494)

In this 1953 essay, "From Fright to Frankenstein" (reprinted in 1976), one can detect the bruised feelings of a man who has found himself caught up in another's "development project."

Some years later Shaw shows himself to have found some humor in his misplaced faith in the Rapid Selector and to be able to apply what he learned from that experience to documentation in general. At a seminar in 1958, following a review of equipment and techniques for information handling, he makes the point that while the machine could scan 100,000 items in four minutes, that number constituted only one year of the *Bibliography of Agriculture*. If one needed to search ten-year runs, one could do only about eight searches in a working day:

> If I do say it myself, the Rapid Selector was a wonderful machine. It was cute, the first one which ever did such wonderful things, and still I could only dig the answers to eight questions from the ten year run of the *Bibliography* [*of Agriculture*] in a day's work. And if any reference librarian couldn't do better than that, one of us would have to go, and it wouldn't cost $100,000 to replace us either. This is the sort of arithmetic you have to learn to apply in this game. The ability to run fast is not enough. (Documentation seminar, 1958, p. 28).

Because he so frequently cautioned librarians against blind faith in machines, he was often accused of being a Luddite. The most famous example occurred in an article by Jesse Shera, "Beyond 1984," published in the official journal of the American Library Association (ALA) in 1967. In it Shera quotes from Shaw's fourteen-year-old "From Fright to Frankenstein" essay, taunts him with the failed Rapid Selector (abandoned by Shaw in 1952), and accuses him of "triviality, error, and even charlatanry" (Shera, 1967a, p. 35). Shaw was so outraged he threatened to sue (Shaw, 1967a). In a rebuttal letter to the *ALA Bulletin* Shaw recites current uses of machines in the University of Hawaii library, which he directed at the time, and succinctly states his position once more: "It is just as stupid to hate machines as it is to love them" (Shaw, 1967b). Shaw never undervalued the usefulness of automation; in fact, in 1961, he urged research so that the National Institutes of Health could experiment with new methods of providing "medical intelligence," including the use of digital computers and other electronic equipment, although he advised that thorough systems analysis and investment in human intelligence be given priority (Shaw, 1961).

The 1967 Shaw-Shera spat did not come out of the blue. There had been friction between them since the early 1950s, although they had been friends for many years, at least according to Shera (1967b). A number of conjectures can be made as to the cause of the friction, but the one that may be most pertinent to this discussion is the suspicion that simple professional competitiveness may have been the culprit (Tefko Saracevic, personal communication, September 1998). Shera was an early advocate of the use of machines for information handling (Shera, 1936). When he worked for the Scripps Foundation for Population Research, he became adept at using tabulating machines (Wright, 1988, pp. 11–12). But it was Shaw who had the dubious pleasure of testing the first electronic bibliographic machine, and it was Shaw who landed a lucrative grant in 1957 from the Council on Library Resources to produce the multivolume *State of the Library Art*. He then rubbed salt in Shera's wounds by attacking the machine that at last emerged from Shera's Center for Documentation and Communication Research at Western Reserve University (Documentation seminar, 1958, pp. 23–24; Shaw, 1963). This unfortunate conflict would be relegated to the realm of old gossip were it not for the fact that Shaw and Shera were major figures who, at least in the early days, were at home in the worlds of both librarians and documentalists. Had they combined forces, they might have reconciled differences between the two groups and perhaps speeded the development of automated retrieval.

Shaw, unlike Shera, seems not to have been much involved in the librarian versus documentalist debate, perhaps because he did not devote as much time to thinking about professional education issues, which had much to do with the disagreements (Williams, 1997). It may be that Shaw perceived the real split to be between the people who were devoted to the machines for the sake of the machine and those who saw the machines merely as tools in the provision of information service. Vannevar Bush, for example, was a visionary and a brilliant engineer, but he had no understanding of the organization of knowledge and little real sympathy for the social function of libraries. Shaw had no use for people who worked on creating new indexing schemes in the abstract, without reference to real collections of information (Shaw, 1963, p. 410). Nevertheless, even though he felt that good indexing depended on human intelligence, he supported doctoral work on automatic indexing (Susan Artandi's dissertation, 1963). In an article in *Science* he suggested that those newly converted to documentation lacked the user perspective and library service application. Here as elsewhere Shaw insisted that one should study information needs from the user's point of view and to think in terms of the total system of scholarly communication (Shaw, 1957; 1962; 1971). A reading of both his 1962 and 1963 *Science* articles today might lead one to conclude that he would have had no trouble seeing how the perspectives and skills of librarians, documentalists, and information scientists could be integrated for the benefit of users.

Conclusion

Shaw's gamble on the Rapid Selector was not a total loss. As Mohrhardt suggests, "The project was as valuable in pointing out what could not be done efficiently with machines as it was in demonstrating the uses of nonbook storage devices" (1957, p. 76). Because of the enormous interest in machine applications at the time, the experiment garnered a great deal of attention and gave Shaw a platform from which he could expound his views. While he was not inclined to blame himself publicly for any of the Selector's failings, he did attempt to prevent others from falling into similar traps. When he warned against accepting machine solutions without adequate preparatory systems analysis, he was implicitly confessing that he had not practiced what he preached. He may never have admitted his shortcomings in the area of representation of knowledge, but at least he recognized and proclaimed consistently the primacy of the intellectual effort required to make the content of scientific literature accessible. He himself did not have the type of mind or the patience to address this aspect of the information problem, nor did he have much tolerance for those who took to it as an abstract exercise. But, writing in journals such as *Science*, he reached an audience that stood to gain from improved access to scientific information, and he explained and promoted the role of librarians and documentalists in that process.

Shaw died in 1972 and thus did not have the opportunity to see the early machine-assisted bibliographic systems evolve into the sophisticated information retrieval of today. It is tempting to speculate that had he lived long enough he would have been among the first to test the efficiency of online searching against manual methods. The saga of Shaw and the Rapid Selector has taught us several lessons: to understand better the interlocking needs and purposes of information users, providers, and systems designers, and to evaluate new technology from that perspective; to avoid confusing tools with systems; and to stay off bandwagons until we know whether they will get us to where we want to go.

References

Artandi, S. S. (1963). *Book indexing by computer.* Unpublished doctoral dissertation, Rutgers University, New Brunswick, NJ.

Bagg, T. C., & Stevens, M. E. (1961). *Information selection systems: Retrieving replica copies: A state of the art report* (National Bureau of Standards Technical Note 157). Washington, DC: Government Printing Office.

Buckland, M. K. (1992). Emanuel Goldberg, electronic document retrieval, and Vannevar Bush's Memex. *Journal of the American Society for Information Science, 43*, 284–294.

Burke, C. (1991). A practical view of Memex: The career of the Rapid Selector. In J. M. Nyce & P. Kahn (Eds.), *From Memex to Hypertext: Vannevar Bush and the mind's machine* (pp. 145–164). San Diego: Academic Press.

Burke, C. (1994). *Information and secrecy: Vannevar Bush, Ultra, and the other Memex.* Metuchen, NJ: Scarecrow.

Documentation seminar, Montreal (1958). New methods and techniques for the communication of knowledge, directed by Ralph R. Shaw, January 20–21, 1958, McGill University. *Proceedings of the Documentation Seminar.* Ottawa, Ontario: Canadian Library Association.

Farkas-Conn, I. S. (1990). *From documentation to information science: The beginnings and early development of the American Documentation Institute–American Society for Information Science*. Westport, CT: Greenwood.

Hines, T. C. (1975). Shaw and the machine. In N. D. Stevens (Ed.), *Essays for Ralph Shaw* (pp. 6–14). Metuchen, NJ: Scarecrow.

Jahoda, G. (1961). Electronic searching. In R. R. Shaw (Ed.), *The state of the library art* (Vol. 4, part 4, pp. 137–317). New Brunswick, NJ: Graduate School of Library Service, Rutgers University.

Miles, W. D. (1982). *A history of the National Library of Medicine: The nation's treasury of medical knowledge.* (NIH Publication No. 82-1904). Washington, DC: U.S. Government Printing Office.

Mohrhardt, F. E. (1957). The library of the United States Department of Agriculture. *Library Quarterly, 27,* 61–82.

Nyce, J. M., & Kahn, P. (1991). *From Memex to hypertext: Vannevar Bush and the mind's machine*. San Diego, CA: Academic Press.

Olivieri, B. L., & Forbes, J. B. (1969). Bibliography of agriculture. In A. Kent & H. Lancour (Eds.), *Encyclopedia of library and information science,* Vol. 2. New York: Marcel Dekker.

Perry, J. W., Kent, A., & Berry, M. M. (1956). *Machine literature searching.* New York: Interscience Publishers.

Shaw, R. R. (1939). Gary charging system. *Library Journal, 64*, 198–199.

Shaw, R. R. (1946). The Publication Board. *College and Research Libraries*, *7,* 105–108.

Shaw, R. R. (1949a). Photoelectric librarian. *Electronics, 22*(9), 122, 158, 160, 164, 166.

Shaw, R. R. (1949b). The Rapid Selector. *Journal of Documentation, 5*(3), 164–171.

Shaw, R. R. (1950). High speed intermittent camera. *American Documentation, 1*, 194–196.

Shaw, R. R. (1951a). Machines and the bibliographical problems of the twentieth century. In *Bibliography in an age of science* (pp. 37–71). Urbana, IL: University of Illinois.

Shaw, R. R. (1951b). Management, machines, and the bibliographic problems of the twentieth century. In J. H. Shera & M. E. Egan (Eds.), *Bibliographic organization; Papers presented before the fifteenth annual conference of the Graduate Library School, July 24–29, 1950* (pp. 200–225). Chicago: University of Chicago Press.

Shaw, R. R. (1953). *The use of photography for clerical routines: A report to the American Council of Learned Societies.* Washington, DC: American Council of Learned Societies.

Shaw, R. R. (1954). Introduction (issue on scientific management in libraries). *Library Trends, 2*, 359–360.

Shaw, R. R. (1957). Documentation: Complete cycle of information service. *College and Research Libraries, 18*, 452–454.

Shaw, R. R. (1958). Mechanical storage, handling, retrieval and supply of information. *Libri, 8*, 1–48.

Shaw, R. R. (1961). *A medical intelligence program for the National Institutes of Health.* New Brunswick, NJ: Rutgers University Graduate School of Library Service.

Shaw, R. R. (1962). Documentation and the individual. *Science, 137*, 409–411.

Shaw, R. R. (1963). Information retrieval. *Science, 140*, 606–609.

Shaw, R. R. (1967a). Letter from Shaw to Mary Gaver, dated February 12, 1966 [actually 1967], ALA Director's Subject Files, Record Series 2/4/6, American Library Association Archives, University of Illinois Archives, Urbana, IL.

Shaw, R. R. (1967b). Beyond 1984—a reply. *ALA Bulletin, 61*, 231–232.

Shaw, R. R. (1971). *Pilot study on the use of scientific literature by scientists*. Metuchen, NJ: Scarecrow Reprint Corporation.

Shaw, R. R. (1976). From fright to Frankenstein. In D. J. Ellsworth & N. D. Stevens (Eds.), *Landmarks of library literature: 1876–1976* (pp. 487–495). Metuchen, NJ: Scarecrow.

Shera, J. H. (1936). The college library of the future. *ALA Bulletin, 30*, 494–501.

Shera, J. H. (1967a). Beyond 1984: What's past is prologue. *ALA Bulletin, 61*, 35–47.

Shera, J. H. (1967b). Letter from Shera to Ruth Warncke, dated January 30, 1967, ALA Director's Subject Files, Record Series 2/4/6, American Library Association Archives, University of Illinois Archives, Urbana, IL.

Stevens, N. D. (1978). Shaw, Ralph Robert (1907–1972). In B. S. Wynar (Ed.), *Dictionary of American library biography* (pp. 476–481). Littleton, CO: Libraries Unlimited.

Turner, I. B. (1983). Ralph Shaw. In W. A. Wiegand (Ed.), *Leaders in American academic librarianship: 1925–1975* (pp. 289–319). Pittsburgh, PA: Beta Phi Mu (distributed by the American Library Association).

Williams, R. V. (1997). The documentation and special libraries movements in the United States, 1910–1960. *Journal of the American Society for Information Science, 48*, 775–781.

Wright, H. C. (1988). *Jesse Shera, librarianship, and information science.* Provo, UT: School of Library and Information Sciences, Brigham Young University.

Zachary, G. P. (1997). *Endless frontier: Vannevar Bush, engineer of the American century*. New York: Free Press.

The Information Wars: Two Cultures and the Conflict in Information Retrieval, 1945–1999

Mark D. Bowles

Abstract

In the generation after World War II a wide perception of an information crisis plagued all scientific professions. This crisis was an "information explosion" that scientists confronted from exponentially increasing numbers of publications. One significant result of this crisis was the emergence of the "information wars"—the professional battle between scientists (documentalists) and humanists (librarians) over information retrieval.

Scientists often blamed an unresponsive library community for failing to develop new techniques to ease the burdens that resulted from too much information. They believed card catalogs were too slow and inefficient, and they desired new automated systems for information retrieval. Librarians often resisted experimenting with these new computing machines because of their expense and technical complexity. As a result scientists began looking elsewhere for bold approaches to solve the information crisis.

The solution that scientists favored came from a relatively new professional group called documentalists. Typically emerging from a scientific background themselves, documentalists began using new punched-card computing machines to facilitate information retrieval for scientific needs. Documentalists believed that their profession represented the future of information retrieval and not that of the antiquated, humanistic librarian.

While the information wars have dominated the information professions over the past half century, as the millennium approaches, the two cultures of information retrieval are now becoming one. With the technological battleground shifted from the scientists-only realm of punched-card machines to the more inclusive and inexpensive technology of the personal computer and the Internet, librarians are emerging once again as the primary gatekeepers of knowledge.

World War II transformed the scientific discipline. Never before in such a dramatic, large-scale, and public way did the results of scientific activity play such an important role in shaping the outcome of world developments. Scientists themselves became national heroes as the nation's strength came to be determined equally by military might *and* by scientific capability. Yet even though some scientific communities seemed to "wear the tunic of Superman" and stand "in the spotlight of a thousand suns," a significant problem reaching crisis proportions plagued all scientific disciplines (Kevles, 1987).

This crisis was an information crisis—a problem of too much information that scientists confronted in the form of exponentially increasing numbers of books, journals, and conference papers (Bowles, 1999). This overload threatened to burden individual researchers with so much data that they feared they would spend all their time quietly reading to keep up with their colleagues. As a result they would be left unable to advance their own ideas, thus ending or curtailing the future progress of science.

The information crisis was one of the most significant intellectual concerns of the twentieth century. I believe this story is important to understand not only because of its central place within the scientific discourse of its time but also because of the conflict it initiated over information retrieval. This conflict centered on the library as a professional battleground between librarians and a relatively new professional group called documentalists. At stake was which professional group would control the future of scientific information.

This conflict was most dramatically played out at Western Reserve University, now Case Western Reserve University, in Cleveland, Ohio. This institution housed one of the nation's leading library schools and documentation centers—the Center for Documentation and Communication Research (CDCR). The CDCR was often referred to as the "best known" of the academic information centers and the "world's most advanced information retrieval system." (On CDCR's significance see "U.S. Organizations," 1961; "Western Reserve Up-Dates," 1961; Kane, 1958; "Scientists Use," 1960; "Take-Off," 1963.) While in many ways this was a unique institution, the attitudes and beliefs held by its documentalists and librarians were a microcosm of a wider professional conflict. Jesse Shera, the dean of the library school (and ironically the driving force behind the creation of the CDCR), described the interrelationship between the information crisis, the emergence of the documentalists, and the threat they posed to the library profession in a 1960 annual report:

> To say that American librarianship today faces the most critical test of its brief history is not rhetorical exaggeration. The growth and increasing complexity of recorded knowledge has not only taken traditional library methods beyond any limits that they were originally designed to meet, but also it has brought into being a rival group who call themselves documentalists, information specialists, or some other name which seems to avoid the use of the term librarian. Thus has been created a schism within the profession that seriously threatens its unity, and that can result only in disastrous consequences to both approaches to the library problem. (Shera, 1960)

Other representative comments reveal the broad awareness of this conflict among the information professions. In 1956 Neal Harlow, a University of British Columbia librarian, wrote, "There has been such a revolution in bibliographic needs that our professional usefulness is being severely challenged." In 1963 a documentalist and manager of IBM's technical information center reported, "The IBM . . . Information Center was born from management's concern that the libraries serving its technical and professional personnel were not geared to the speed and complexity of present requirements (White, 1963). Also in 1963 William S. Budington, a president-elect of the Special Libraries Association wrote that there was a growing perception that scientists and engineers "were required to give birth and nurse the necessary gadgets" to solve the information crisis, and not the librarians with their "creaky procedures." In 1972 Marilyn Gell, a Virginia public librarian, wrote a modern fable called "The Passing of the Unicorn" in which this once proud animal (the unicorn representing the librarian) was threatened with extinction by "no-horners" (the modern beast representing the documentalists) who sought to "computerize its wisdom."

These examples (from a university librarian, a documentalist, a special librarian, and a public librarian) serve as contextual indicators that this professional conflict was a national phenomenon and not localized to a specific institution or group of practitioners. The preceding statements also reveal that a technological debate was central to this conflict, as documentalists wanted to use new punch-card computing machines to solve the information crisis, while a majority of librarians seemed to resist the new devices.

Why was there a professional conflict between documentalists and librarians? One main reason was that librarians were typically humanists and documentalists were scientists. The difference in professional background is not a trivial one. Many have referred to the science–humanities distinction as one of the most significant intellectual chasms of the twentieth century. This phenomenon was first brought to widespread attention in the 1950s by British scientist-novelist C. P. Snow; he described the split as the "two cultures." Snow argued that all Western intellectual activity was splitting into two polar groups. Humanists or literary intellectuals were at one pole and scientists were at the other. Between them was a "gulf of mutual incomprehension—sometimes . . . hostility and dislike, but most of all a lack of understanding" (Snow, 1961).

Historian of science Alan Rocke (1998) recently commented that the conflict between scientists and humanists is a cultural divide that continues to the present. Rocke claimed that one result was something called "the science wars"—the debate over how scientists and humanists understood the making of science. Using similar terminology, I argue that when the history of information during the last half of the *twentieth* century is analyzed, it is a story best characterized as the "information wars." Librarians, with their strong background as humanists, lost part of their identity, power, and profession in their battle against the documentalists and scientists. They lost this battle because of such cultural obstacles as the privileged position of the sciences in relationship to the humanities and because they resisted the coming of the computer to the library.

Calvin Mooers, who coined the term *information retrieval*, represents one of the best examples of this conflict. Mooers, inventor of the Zatocoding system for information retrieval, not only identified but was also a key participant in the conflict between the two cultures of information retrieval. As a mathematician his disdain for the capabilities of the humanistic librarians was often apparent. In a private letter to library school dean Jesse Shera in 1957, Mooers expressed his concern over what he saw was the emergence of "two cultures" at the American Documentation Institute conferences. On the one side were the people who were building the "machines of the future," and on the other side were the librarians (Mooers, 1957). What he thought was unfortunate was that the machine people could "peer into the mysteries of the library" and understand and improve upon the activities found within. Yet the librarians were unable to do the same with the machines. Mooers said librarians found his machines "repugnant," his devices "antagonized" them, and the librarians were left "baffled" (Corbitt, 1993). Charlotte Mooers shared her husband's perceptions of librarians. She recently recalled that most of the people to whom Calvin explained the Zatocoding system did not understand it, but she confessed "quite frankly the people who didn't understand it *the most* were the librarians" (personal communication, 21 May 1998). It was true that Mooers wanted to develop a machine to replace the librarian in the search for information. The librarians naturally were repulsed by this idea, but Mooers joked that librarians "took offense against the idea even though they weren't able to fully formulate why they were offended by it" (Corbitt, 1993).

Neither librarian nor documentalist emerged from this professional warfare the victor, and both suffered serious setbacks to their disciplines. For example, neither the renowned library school nor the documentation center at Western Reserve University exist today. However, as our millennium ends, the two unique cultures represented by the documentalists and the librarians are now becoming one, as the technological battleground has shifted from the scientists-only realm of punch-card machines to the more inclusive and inexpensive technology of the personal computer and the Internet. The following is the story of the scientific information crisis, the resulting information warfare, the use of weapons of automation, and an emerging information détente.

A Scientific Information Crisis

In their anthropological study of the life inside a scientific research center, Bruno Latour and Steve Woolgar examined the daily existence of a scientist in the laboratory (1986). While they were not surprised to learn that scientists read published material, they were unprepared to discover the "central prominence of documents" and the "vast body of literature emanat[ing] from within" the laboratory. They found that the scientists were "compulsive and almost manic writers" and that the laboratory surrounding them was a "hive of writing activity."

The scientists' written reports became the central product of their research, as the entire working day seemed to revolve around the production of written material. Every discussion between the scientists, no matter how brief, always focused on the published literature including informal discussions, telephone conversations, and official presentations. Latour and Woolgar were perplexed in the confrontation with this "strange tribe" and their "omnipresence of literature." By 1945 this literature overload threatened to strangle future scientific progress and became a major concern for the scientific community.

Because of the centrality of documents to the scientific profession, any threat to the information retrieval and dissemination system was regarded as a significant problem. As the publication of journal articles, books, and conference papers began to overwhelm the scientists, they came to the conclusion that they were experiencing an information crisis. Like any other finding in the laboratory, scientists used their written output to convince others of this assessment. The immediate goal of this persuasion was to stimulate work directed at finding a solution to the problem. These concerns quickly spread throughout all scholarly disciplines—particularly engineering.

Why did this problem appear to emerge so suddenly after 1945? There were three key reasons. The first reason was World War II. As one observer wrote, the war "wrecked" the scientific communication system (Bernal, 1944–45). Indeed, the scientific mobilization and effort for the war was directed single-mindedly toward military success. As a result scientists had little time to publish their work, and much of that work was classified as secret. Thus, when the war was over and the government lifted the secrecy ban, a large body of research was made available through publications.

Second, the Cold War played a key role in heightening the sense of an information crisis. The dramatic and visible success of the Soviet *Sputnik* satellite in 1957 demonstrated the real possibility that American science and engineering were falling behind that of their Communist counterparts. Furthermore, evidence of a vast centralized information network at the Soviet All-Union Institute of Scientific and Technical Information (VINITI), greatly concerned U.S. scientists. Reports indicated that this institute employed twenty thousand abstractors and translators to effectively disseminate information to Russian scientists and engineers. Assisting their work in this was a rumored massive punched-card machine feared to be the computer equivalent of *Sputnik*. Through this centralized information service it appeared as if the Soviets might have solved the information crisis itself.

Finally, the information crisis that emerged after 1945 was in part the result of the natural perception by contemporaries that scientific growth was out of control. As historian of science Derek J. de Solla Price (1963) observed, exponential growth was such a central feature of scientific activity that it is "the fundamental law of any analysis of science." The result was his often-quoted, astounding fact about the scientific discipline: "80 to 90 percent of all the scientists that have ever lived are alive now." This statement was as true in 1660 or 1945 as it is today. However, when this natural state of exponential growth was coupled with the circumstances surrounding the end of World War II and the emergence of the Cold War, this situation became a true "crisis."

Chemists were particularly concerned. For many years the chemical profession knew the value of organizing its published information. *Chemical Abstracts* had long provided summaries of the world's chemical literature and, even today, boasts the largest resource on chemical information. But the growing amount of published literature threatened to overwhelm the editors of this abstracting journal. In 1949 editor E. J. Crane (1949) examined the publication increase in his journal because he thought this would be a "reasonably good yardstick" to measure the increase in research in other fields. He made the following findings:

- The *Journal of the American Chemical Society* increased its number of articles by 63 percent from 1947 to 1948.
- *Industrial and Chemical Engineering* increased 45 percent in 1948.
- *Physical Review* had a backlog of over eight hundred papers waiting to be published.
- The *Journal of Biological Chemistry* increased 63 percent from 1947 to 1948.
- *Chemical Abstracts* planned to increase its coverage of the literature by 21.1 percent

Herein lay the heart of the crisis. If most scientific journals were increasing by as much as 60 percent in a given year, and *Chemical Abstracts* planned only a 21 percent increase, then how many significant articles would be overlooked and ignored? Crane concluded, "Chemical publication is literally booming. I have never seen anything like it."

Other chemists agreed with Crane and were equally concerned. For example, one chemist (Richardson, 1951) claimed that there was "too much current literature on chemistry . . . and it is not properly organized." A biochemist (Archibald, 1952) argued that the "volume of literature . . . is increasing so rapidly . . . that lack of appreciation of what has been achieved by others is limiting markedly our scientific productivity." The editor of *Chemical & Engineering News* argued that his editorial work was more "hectic" than his predecessors, claiming that from 1929 to 1950 the journal increased in size by 760 percent (Murphy, 1951).

The scientific information crisis was not confined to the chemical discipline; it was also a concern of many leading scientists from 1945 to 1963. For example, an engineer at the Stanford Research Institute described the "technical literature problem," a biologist at the American Institute of Biological Sciences identified the "critical problem of research publication," the president of the American Society for Metals complained of the "literature jungle," the director of the National Science Foundation described the "information problem," and the director of the Oak Ridge National Laboratory specifically called all of these problems "the information crisis." The engineer was Charles P. Bourne. The biologist was John A. Behnke. The American Society of Metals president was Walter Crafts. The NSF director was Burton W. Adkinson. The Oak Ridge National Laboratory director was Alvin M. Weinberg.

Implicit in this concern over the information crisis was an attack upon the library. The best-known spokesman of this attack was Vannevar Bush, the main architect of science policy during World War II. Bush was actually the first to define the information crisis in the postwar era. As J. C. R. Licklider (1965) stated, "Vannevar Bush . . . may be said to have opened the current campaign on the 'information problem.' " In 1945 Bush wrote his now legendary article in *Atlantic Monthly* called "As We May Think." While many scholars, such as

Michael Buckland (1992) and W. Boyd Rayward (1994), now rightly argue that Bush's ideas were not nearly as novel as once believed in terms of his Memex, the article was important for crystallizing the concerns of the information crisis for a wider scientific and technical audience. Bush said (1945), "The difficulty seems to me not so much that we publish unduly . . . but rather that publication has been extended far beyond our present ability to make real use of the record." Ten years later Bush (1955) published the article "For Man to Know," saying, in what became a popular and often-quoted phrase, "Science may become bogged down in its own product, inhibited like a colony of bacteria by its own exudations." The product of science to which Bush referred was the publication.

Bush consistently tied his concerns about the information crisis to an attack on the library. He (1953) believed that the library was unable to meet the needs of scientists and felt that librarians were inadequate guides to the relevant literature. Bush shifted blame away from scientists by saying that they were not publishing too much, but that librarians were not managing their output effectively (1945). Colin Burke (1994, p. 119) has suggested recently that "Bush wanted a fundamental reform of the library to make it conform to the concepts of the new scientists and engineers," using machines to allow scientists to simply "bypass the library." These criticisms of the library and of the librarians' reluctance to find a solution became widespread.

What further intensified the information crisis was a bit of interesting irony. As scientists were leveling attacks against the inefficiencies of the library, the library was itself becoming of increasing importance to the scientist. And yet, scientists believed that the library was an overwhelmed and outdated institution that was failing to cope with this outpouring of information. To the rescue came not the librarian, but a new information professional—the documentalist with a commitment to the automation of information for the specialist.

Information Warfare: Conflict in the Library

In *When Old Technologies Were New,* historian Carolyn Marvin examined the early history of electric media (electric light and the telephone) and postulated that its history was less a story of the evolution of instruments and more about social groups negotiating power. Issues surrounding these groups concerned who was "inside" (the professional electricians) and who was "outside" (the public), who had authority and who had none. Marvin observed (1988) that "new media intrude on these negotiations by providing new platforms on which old groups confront one another." This media could "change the perceived effectiveness" of one of these competing groups. For example, a group that possessed the latest technical know-how could define themselves as experts and use this status as a claim to authority. This type of social conflict and power negotiation also became central to the early history of the computer and of information processing. A new professional group (the documentalists) threatened to wrest the control of information away from the traditional group in power (the librarians) by using a new electronic media (the computer).

During the late 1950s the conflict over which professional group was best suited to control information was portrayed on the silver screen by two of Hollywood's most popular stars, Katharine Hepburn and Spencer Tracy. Their 1957 film *Desk Set* took place at a fictitious television studio called the Federal Broadcasting Company, where Hepburn worked as a reference librarian. Tracy played a "methods engineer," one of the leading experts on "electronic brains" in the country, who had been hired by the broadcasting company to automate and replace the jobs of the reference librarians. When Hepburn's character first saw the electronic brain, she said it was "frightening" and was a "monster machine." When she learned of Tracy's character's professional background, she immediately reached for a cigarette and whispered to another librarian, "I only smoke when there is a crisis. . . . He is an engineer."

Despite the happily-ever-after Hollywood ending with the librarian and the engineer falling in love, the film raised a number of important issues concerning the future of the information professional in an increasingly computerized society: a fundamental conflict between the humanistic librarian and the scientific information professional; the fight for control of the library; and the librarians' fear of automation. These themes were fictionalized versions of what was actually occurring in the conflict between the documentalists and librarians.

Documentalists were not new to the post-1945 period. Irene Farkas-Conn, Robert V. Williams, and W. Boyd Rayward have expertly analyzed the history of documentation that extends back to the turn of the twentieth century. Founders such as Paul Otlet in Belgium and Watson Davis in America made significant advances in organizing information (Rayward, 1997, 1975). One of the first Americans to become interested in the documentation field was Watson Davis. Davis began his career as a civil engineer at the National Bureau of Standards while simultaneously becoming one of the first journalists to report on scientific developments. Working for the Washington *Times-Herald* in the 1920s alerted him to the significance of scientific communication, and he became interested in documentation as a way to solve the difficulties associated with a growing amount of information. In 1933 he headed the Science Service in Washington, D.C., which attempted to popularize science and gain funding for it. In 1937 Davis called together thirty-five of his colleagues to meet at the National Academy of Sciences. According to his daughter Charlotte Mooers, Davis proposed the formation of the American Documentation Institute (ADI) to research new ways of disseminating scientific information to a wider audience through microfilm technology (personal communication, 21 May 1998). He said that he liked using the term "documentation" because it had an international reputation (owing to Otlet) and was inclusive of all forms of intellectual activity. This included librarians and humanists and was "not specifically limited to the fields of the physical and natural sciences (1935)."

By the post–World War II period those who called themselves documentalists narrowed their customer base primarily to serve the sciences. Scientists were the ones most vocal about the information crisis, and during the heightening of the Cold War, government contracts for scientific activities were flowing quickly. Colin Burke wrote, "The achievements in handling tons of documents during World War II allowed the first documentalists to seize the new opportunities of the Cold War and to gain the funding they had begged for during the 1930s" (1994, p. 112). Thus the documentalists wanted to tame the information crisis by becoming the main professional group for controlling information. They based their claim on the growing perception that traditional librarians were not responding to the information needs of scientists. The documentalists seized this opportunity, regardless of the veracity of this perception.

If this meant an intense struggle with librarians, then documentalists were ready for the fight. For example, examining a passage from Farkas-Conn's history of the documentation movement, we can easily see contentious warlike imagery. She wrote, "Like soldiers on the front, [documentalists] had to be preoccupied with the battle, of winning a skirmish; only a few could think of the grand strategy of winning the war, let alone consider the even greater overall plans, the larger societal concerns of establishing peace among the warring parties" (1990, p. 196). Her use of warlike imagery in these descriptions accurately represents the belligerent relationship between these two professions. Farkas-Conn also noted that by 1952 ADI experienced a transition from a new "vital force [which] came from the people who found that traditional library and bibliographic methods were inadequate for the management of scientific and technical information" (1990, pp. 183–184, 186). The real difficulty was that the documentalists and librarians had vastly different backgrounds and outlooks about how to manage scientific information and what to do about science in crisis. Let us examine four of these differences.

First, the importance of the documentalists' and librarians' backgrounds has already been suggested. The documentalist typically emerged from a science or engineering school with an interest in information. Allen Kent, one of the CDCR directors, commented that not only did most documentalists have scientific backgrounds but many were specialized as chemists. He reasoned that chemists had advantages over professionals in other fields because of their understanding of molecular structure notation systems. He believed that this enabled them to be one step ahead of others interested in information searching and especially those like librarians who sought to use outdated alphabetical indexing systems (personal communication, 20 March 1998).

By contrast, librarians had backgrounds in the humanities. This was a career path that documentalists believed librarians turned to because they were "self-consciously inadequate in science" (Shera, 1967). Whether or not librarians were inadequate in science is to be debated, but psychological and vocational profiles given to librarians in the 1950s indicate a strong interest in humanistic endeavors. In 1952 Alice I. Bryan (pp. 29, 31, 35, 43) studied the professional profiles of 2,400 public librarians and concluded that their interests were most comparable to artists, musicians, and writers. Of the 2,395 librarians surveyed, 92 percent were female. The most "significant feature of the age distribution" was what she called the "middle-age bulge," with the median age being 42.3 years. Most of these women, about 75 percent, were unmarried. Bryan gave the female librarians the GAMIN personality test. Her findings were that they were submissive, lacked self-confidence, and had feelings of inferiority. A similar study in 1957 (Douglass) found librarians far more attracted to aesthetics than to science or technology. Such profiles illustrate the different professional and personal interests between documentalists and librarians. Most important, this difference was viewed as a weakness by documentalist-scientists.

A second key difference between documentalists and librarians was their institutional base. The "scientific information center" was the institutional home of the documentation movement. By 1961 there were approximately 221 of these centers in the United States, supported and funded by government, industry, or academia and employing more than 6,000 personnel (Simpson, 1962). Many scientists rejected the library because they did not want an information repository that simply stored data "as a warehouse is used for storing bales of cotton." Instead they wanted a place where data were correlated and distributed to users with specific needs. It was here that many scientists believed the "libraries have failed and failed badly." Many scientists considered the librarians' classification scheme too vague, and some documentalists feared the library would become science's Waterloo.

Third, the customers for these institutional centers also were very different. While librarianship by very definition was inclusive of all people who sought information, the documentalists' service was much more exclusive. A course brochure from the Center for Documentation and Communication Research read: "Documentalists serve, in particular, the interests of research and scholarship; they are not concerned with popular, recreational, or lay interests" (Program for Documentation, 1961). Allen Kent (1961) argued that, in contrast, librarianship was a "passive activity that [could] cope with the general needs of adults and children, but not with the active industrial and governmental requirements of a modern society." In short, the documentalists wanted to deliver to an elite customer base the ability to control and make sense of information in a way that was unavailable to the public. Librarians, on the other hand, were concerned not just with scientists' needs but also with the needs of humanist scholars and with even what some regarded as the "lowest" form of information request—popular reading from a lay audience.

Finally, these two professions had differing notions of what constituted information. Librarians were the defenders of the book as the basic unit of information, while documentalists believed the data contained in books, research papers, technical reports, and governmental studies were the unit of information. This concept of information was implicit in their name—documentalists. It was the document or specific fact that was privileged, not its surrounding context—the book or journal itself. It was impossible, they believed, to effectively communicate the essential information contained within a book through only a title and author on a 3×5-inch library card. Librarians responded by saying that it was the "enemies of libraries" who "tend[ed] to believe that the only things that matter in any book are discrete paragraphs of information." Librarians believed that without the context of the entire book, the individual fact could mean nothing (Crawford & Gorman, 1995). But the documentalists ignored these concerns and continued to attack the book because it lagged three to ten years behind the most up-to-date information (Bree, 1963). For example, Lowell A. Martin (1955) argued, "This is an age of rapid communication; the book by its very nature is slow in bringing its message." Documentalists did not want to do away with the book; rather they wanted the book to exist as a "reservoir" of knowledge.

In each of these oppositions (science–humanities, information center–library, elite–public, and fact–book), the documentalists occupied the culturally privileged position, thus solidifying their source of power. It was in this way that such a young "outside" profession was able to erode much of the status of a long-established "inside" profession to become the new "insider." To make matters worse, there was one further distinction between librarians and documentalists that probably represented the most important aspect of their professional conflict—technology. Documentalists were zealous advocates of new computer technology. Librarians did not simply prefer the card catalog over the computer; they frequently recoiled in fear from the prospects of automation.

Weapons of Automation: The Librarian and the Fear of Computers

The history and tradition of librarianship is long and distinguished. It extends over two thousand years, back to the Alexandrian Library and Eratosthenes. Throughout this history libraries have often been at the forefront of new and revolutionary technology. Historian John Higham (1979) wrote, "We have forgotten how revolutionary a dictionary catalogue of loose cards was when introduced at the Harvard Library in 1861." In 1876 librarians established themselves as a professional organization with the founding of the American Library Association. Libraries grew quickly by becoming an essential organization for meeting the information needs of America's growing industrialization (Harris, 1995). After World War II a few librarians even believed that it would be the library profession that would "take the incentive and attempt to provide leadership" in the search for the solution to the information crisis. Some predicted that in solving this crisis the "library profession [would] make a major contribution" to its own profession and at the same time benefit all society (Egan & Shera, 1949).

But by the 1950s the librarian represented the " 'old guard' in information retrieval" to a majority of library users (Shultz, 1961). No longer could they maintain their position at the forefront of new technology, in large part, because of the expense of computing machines and the technical background required to program them. Few traditional librarians had the opportunity to learn about new mechanized approaches to manipulate information. While the special librarians of the prewar years were advocates of new library mechanisms, even they became more technologically conservative after the war. Farkas-Conn (1990, p. 206) wrote that by the 1950s librarians came to "adamantly oppose automation."

The new automated bibliographic techniques established by the documentalists were presenting professional image problems for librarians. A. J. Goldwyn (1963) questioned, "Will the library of tomorrow need a librarian, or will the librarian be the dodo . . . the technologically displaced ghost." Alan M. Rees (1964a) called the librarian's problem a "personal crisis in the form of the challenge of automation," noting the unnerving fear that the "phantasmagoria of librarian-type robots" would soon displace them (Rees, 1964b). As a result librarianship had a stigma attached to it. No one, it seemed, wanted to be a librarian. The reason that documentalists preferred titles like "manager of technical information" was that "to be tarred" with the name librarian would mean a "loss in salary and status" (Rees, 1964b). So while the scientists faced an information crisis, the librarians were in the midst of an identity crisis. The source of this crisis was their fear of automation. "In countries like the United States, the advent of the 'information age' has provoked major debates about the future of books and libraries and has stimulated wild flights of imagination and fear" (Harris, 1995, p. 294).

Fear was a word frequently used to describe the librarians' reaction to computers. Jesse Shera (1967, p. 749) explained that "fear is especially strong . . . when the innovation [the computer] comes from [outside] the occupational group or subculture." As a result librarianship had a growing antiquated image. Librarians were fearful of the documentalists' "invasion" into their social space wielding a weapon of automation that they could neither combat nor understand.

The so-called "antiquated" librarians did not even trust the language used to describe these machines. *Searching algorithms*, *random access storage*, *Zatocoding*, *zone bits*, and *digital computing* represented an incomprehensible, threatening language and set of ideas for the professional librarian. The computer—or as some librarians referred to it, the "bête noire of the library profession," the "diablus ex machina," and the "Pandora's chest from where all evil swarms"—became a symbol of the librarians' failure to rescue a scientific enterprise in crisis (Shera, 1961).

The documentalists urged librarians not to fear mechanization since the only thing that would be lost was the "drudgery" of the repetitive operations of their work (Bristol, 1952). IBM representative H. S. White (1963b) implored his librarian audience, "Don't be afraid of machine equipment." But the librarians believed that even if the computer could relieve them of "burdensome detail," they would lose control of their profession to the outsiders. Shera revealed that behind all these concerns lay the "fear of loss of professional identity." Concern was generated not only by the machine itself but also by a number of articles proclaiming that machines would soon be in control in the library. For example, R. R. Shaw (1954) titled his article "Will Machines Take Over?" and *Chemical Week* published "Machine Age in the Library" (1954). Shera responded to these articles with one of his own in *Science*, called "Librarians against Machines" (1967).

But it was not necessarily the fear of the mechanical aspects of automation that concerned the librarians. Ralph H. Parker (1965), a documentalist at the University of Missouri, wrote, "When we hear expressions of fear of the machine, what is really meant is that we fear other men's use of it." It would be inevitable, according to Parker, that the computer would replace the menial tasks of the librarian. He was optimistic about the future because the "automation of records in libraries will free librarians, whether they wish it or not, to become truly professional." Phrases like "whether they wish it or not" are important. How could librarians maintain professional status if a group of outsiders dictated the conditions of their professional status?

The results of the librarians' resistance to automation during the 1950s were felt throughout the library profession for the next three decades. In many places across the United States library education did not survive. From 1978 to 1991 fourteen of the most prestigious library schools shut down, including the University of Chicago (the first school to offer a Ph.D. in library science), Columbia University, and Case Western Reserve University (Paris, 1991). This might not seem devastating to nonlibrarians, but imagine what would happen to the engineering profession if MIT and

Stanford shut down their programs. These library school closings represented a very serious problem in the profession. Numerous reasons have been given for the failure of these programs, including lack of funding, a tight job market, academic isolation, complacent library school leaders, and poor quality of the schools. But the failure to understand the technological transition brought on by the computer has been singled out as one of the central reasons that librarianship has suffered (Foster, 1993). In *The Closing of American Library Schools* Ostler, Dahlin, and Willardson (1995) argued that while the nation was changing to an "information society, library school leaders on the whole failed to recognize and adapt in any significant way to this fundamental societal change."

Allen Kent prophetically wrote in 1961 that the "division between librarianship and documentation was not healthy for either." Today documentalists no longer exist. By the early 1960s the term *documentation* itself acquired an old-fashioned image, and in 1963 the profession began to consider a name change, eventually settling on information science (Farkas-Conn, 1990, p. 191). Like the library profession prestigious documentation centers also closed, most notably Kent's own CDCR, which closed its doors in 1971. To announce its passing the university magazine ran just a small announcement, "Documentation Center Absorbed." This hushed closure was buried between news of upcoming campus films for students and an editorial note of when the next issue would appear. How sad that an institute once called a leading information center in the Western world was reduced to a news item that was given no more importance than a local film schedule—all within a span of sixteen years, from 1955 to 1971. The closures of both the library school and CDCR at Case Western Reserve were symbolic of the futility of the information wars. If the information professions could not work together, it was clear that their futures were in jeopardy.

Information Détente: Two Cultures Becomes One

In thinking about the "science wars," Alan Rocke (1998) concluded, "A final suggestion is this: The warriors on both sides of this conflict should calm down, actually read the works of their opponents, and always be intellectually generous to colleagues in different specialties. We are all cultivating the same vineyard." While the larger scientific and humanistic communities have yet to heed this advice, the information professions raised the white flag and have made significant gains at bringing the information wars to an end.

In a note of optimism, Robert V. Williams (1997) wrote that the "fracturing of the information profession" might eventually lead to "greater 'healthiness.' " I agree. Librarians are overcoming their fear of computers, and the two cultures, at least within the realm of information management, may eventually become one. It appears that unlike in the 1950s the library professionals of the 1990s, for the most part, are embracing the new technology. In a decade technologically defined by the personal computer, online databases, and the Internet, librarians are taking an active role and no longer passively resisting technological change.

Some traditional librarians still offer resistance. Just before he died in 1982, Jesse Shera (1983) offered a final warning about the computer. It "must be kept in its proper place as a tool and a slave, or we will become sorcerers' apprentices, with data, data everywhere and not a thought to think." This statement crystallizes the heart of the traditional librarians' fear, concern, and distrust of computers—the fear of servitude to a machine. Early in the 1990s one observer at Michigan State University said that while Internet technology was of interest to librarians, it was also "frightening at the same time" (Charbuck, 1993). In 1992 Charles Robinson, a Baltimore librarian, commenting on the rigid unchanging library profession, stated, "Most of us, quite naturally, will resist the changes that are necessary."

But less traditional librarians know that "virtual, digital libraries are emerging—regardless whether traditional libraries want them to or not" (McClure, Moen, & Ryan, 1994). They believe that librarians should help shape these developments and not simply respond to them. They could either be "dragged kicking, screaming, and whining into a new digitally based information age or they [could] take the lead in making this new information environment better than the last" (McClure et al. 1994, p. 336; Gardner, 1995, p. 15).

Most librarians are now enthusiastically embracing computer technology and using it to increase the status and capabilities of their profession. The *Chicago Tribune* (Swanson, 1995) referred to librarians as some of the "most enthusiastic travelers on the information superhighway." New librarians are increasingly seeing computers not as a threat but as an economic and professional stimulus. Graduates from library schools frequently take jobs in nontraditional library settings with a strong technological emphasis to their work. The result has been a reinvigoration for the library profession.

A librarian at the University of Texas exclaimed, "There's new zip to the stereotypical profile of a librarian" (Murphy, 1997; Schneider, 1996; Thomas, 1995; and Blades, 1994.)

Because of the librarians' acceptance of new technology, even scientists now accept and rely upon the profession of librarianship. At the latest Library of the Future, located at the heart of the Case Western Reserve University campus, humanist and scientific researchers alike are told that the first step for any successful research career is to "get to know the reference librarian in your subject area" (Welcome, 1996; Gopalani, 1997). Understanding the issues surrounding the information crisis in the generation after World War II can assist our planning for other libraries of the future as well as the future of information access. The mistakes made in the past were most notably the absence of the librarian's voice in issues relating to automation and information. We cannot afford that voice to be silenced again, nor can we allow the information wars to claim another victim. The librarian must remain our primary gatekeeper of knowledge.

References

Archibald, R. M. (1952, October 27). Frontiers of biochemistry, *Chemical and Engineering News 30,* 4473–4476.

Bernal, J. D. (1944–1945). Lessons of the war for science. *Reports on Progress in Physics 10,* 418–436.

Blades, J. (1994, December 26). Libraries' place. *Chicago Tribune*, 5, 2.

Bowles, M. D. (1999). *Crisis in the information age? How the information explosion threatened science, democracy, the library, and the human body, 1945–1999*. Doctoral dissertation, Case Western Reserve University, Cleveland.

Bree, R. (1963). Of needles and haystacks, *EURATOM Bulletin*, *2* (19).

Bristol, R. P. (1952, April). Can analysis of information be mechanized? *College and Research Libraries 13*, 131–135.

Bryan, A. I. (1952). *The public librarian*. New York: Columbia University Press.

Buckland, M. K. (1992, May). Emmanuel Goldberg, electronic document retrieval, and Vannevar Bush's Memex. *Journal of the American Society for Information Science,* 284–294.

Budington, W. S. (1963, July–August). Special librarians and the Weinberg Report. *Special Libraries*, 326.

Burke, C. (1994). *Information and secrecy: Vannevar Bush, ultra, and the other Memex*. Metuchen, NJ: Scarecrow Press.

Bush, V. (1945, July). As we may think. *Atlantic Monthly, 176,* 101–108.

Bush, V. (1953). Science and medicine. *Medical Annals of the District of Columbia, 22* (1), 5.

Bush, V. (1955, August). For man to know. *Atlantic Monthly 186*, 32.

Charbuck, D. C. (1993, February 15). Good-bye, Dewey Decimal. *Forbes,* 204–205.

Corbitt, K. D. (1993). Calvin N. & Charlotte D. Mooers interview (22 June), OH 254. University of Minnesota: Charles Babbage Institute.

Crane, E. J. (1949). Scientific publication now very active. *Review of Documentation (FID)*, *16*, 49–53.

Crawford, W., & Gorman, M. (1995). Most data is meaningless outside its local sphere. *Future libraries: Dreams, madness, & reality*. Chicago: American Library Association.

Davis, W. (1935, July 11). Regarding the name: Documentation Institute of Science Service. Calvin N. Mooers Papers, CBI 81, Box 18, Folder 18. University of Minnesota: Charles Babbage Institute.

Douglass, R. (1957). The personality of the librarian. Doctoral dissertation, University of Chicago.

Egan, M. E., & Shera, J. H. (1949). Prolegomena to bibliographic control. *Journal of Cataloging and Classification, 5* (Winter), 19.

Farkas-Conn, I. (1990). *From documentation to information science*. New York: Greenwood Press.

Foster, S. P. (1993). Victimization in library school closing rhetoric: A response to a *Library Quarterly s*ymposium. *Library Quarterly, 63*(2), 199.

Gardner, M. (1995, April 13). Redefining libraries in the age of the Internet. *Christian Science Monitor*, 15.

Gell, M. (1972). The passing of the unicorn. In C. West & E. Katz, (Eds.). *Revolting librarians* (pp. 1–2). San Francisco: Booklegger Press.

Goldwyn, A. J. (1963, July). The advance of automation: Its implications for library education. *Library Journal*, 2640–2643.

Gopalani, R. (1997). Kelvin Smith library upgrades its facilities. *The Observer, 28*(19), 1.

Harlow, N. (1956, May 1). Documentation and the librarian. *Library Journal, 81*(9), 1084.

Harris, M. H. (1995). *History of libraries in the western world* (4th ed.). Metuchen, NJ: Scarecrow Press.

Higham, J. (1979). The matrix of specialization. In A. Oleson & J. Voss (Eds.). *The Organization of knowledge in modern America, 1860–1920*, (p. 14). Baltimore: Johns Hopkins University Press.

Kane, R. W. (1958, January 17). WRU plans world center here for document study. *The Plain Dealer*, 1.

Kent, A. (1961, June). Documentation and communication research or confessions of a Cleveland documentalist on the occasion of the American Library Association Conference in that city. *Wilson Library Bulletin, 36*(6), 1.

Kevles, D. J. (1987). *The Physicists: The history of a scientific community in modern America.* Cambridge, MA: Harvard University Press.

Latour, B., & Woolgar S. (1986). *Laboratory life: The construction of scientific facts.* Princeton, NJ: Princeton University Press.

Licklider, J. C. R. (1965). *Libraries of the future*. Cambridge, MA: MIT Press.

Machine age in the library. (1954, March 13). *Chemical Week, 74*, 74–76.

Martin, L. A. (1955). Implications for the education of librarians. *Library Quarterly, 25*(4), 365–366.

Marvin, C. (1988). *When old technologies were new: Thinking about electric communication in the late 19th century*. New York: Oxford University Press.

McClure, C. R., Moen, W. E., & Ryan, J. (1994). *Libraries and the Internet/NREN: Perspectives, issues, and challenges*. Westport, CT: Mecklermedia.

Mooers, C. N. (1957). Private letter to Jesse Shera (5 January), CBI 81, Box 18, Folder 52. University of Minnesota: Charles Babbage Institute.

Murphy, K. (1997, January 6). Moving from the card catalog to the Internet. *New York Times*, p. C5.

Murphy, W. J. (1951, July 30). The pace now is much faster. *Chemical & Engineering News, 29,* 3077.

Noble, D. F. (1984). *Forces of production: A social history of industrial automation*. Oxford: Oxford University Press.

Ostler, L. J., Dahlin, T. C., & Willardson, J. D. (1995). *The closing of American library schools: Problems and opportunities*. Westport, CT: Greenwood Press.

Paris, M. (1991). Perspectives on the elimination of graduate programs in library and information studies. *Library Quarterly, 61*(3), 259.

Parker, R. H. (1965). The machine and the librarian. *Library Resources & Technical Services, 9*(1), 100.

Price, D. J. De S. (1963). *Little science, big science*. New York: Columbia University Press.

Program for documentation specialists (1961). Center for Documentation and Communication Research, Box 1, File 45A-21 Brochures CDCR. Case Western Reserve University Archives, Cleveland, OH.

Rayward, W. B. (1975). *The universe of information: The work of Paul Otlet for documentation and international organization.* Moscow: VINITI.

Rayward, W. B. (1994, May). Visions of Xanadu: Paul Otlet (1868–1944) and hypertext. *Journal of the American Society for Information Science, 45,* 235–250.

Rayward, W. B. (1997). The origins of information science and the International Institute of Bibliography/International Federation for Information and Documentation (FID). *Journal of the American Society for Information Science, 48*(4), 298.

Rees, A. M. (1964a). The future of the special librarian—bright or bleak. CWRU Archives, Box 45ED2.

Rees, A. M. (1964b). New bottles for old wine: Retrieval and librarianship. *Wilson Library Bulletin, 39*(5), 773.

Richardson, A. S. (1951, May 28). Foundations of chemistry. *Chemical & Engineering News, 29,* 2134–2136.

Robinson, C. (1992, March 15). The public library vanishes. *Library Journal*, 51.

Rocke, A. (1998, November). The science wars. Unpublished lecture. Presented at Case Western Reserve University.

Schneider, K. G. (1996, January). Community networks: New frontier, old values. *American Libraries*, 96.

Schultz, C. K. (1961). Information retrieval: Some implications for smaller libraries. *Wilson Library Bulletin, 35*(10), 784.

Scientists use new computers to index technical reports. (1960, April 7). *Wall Street Journal*.

Shaw, R. R. (1954, April 15). Will machines take over. *Library Journal 76*, 722–724.

Shera, J. (1960). *School of Library Science annual report (1960–61).* Cleveland, OH: Western Reserve University.

Shera, J. (1961). Automation without fear. *ALA Bulletin, 55*(9), 787.

Shera, J. (1967, May 12). Librarians against machines. *Science, 156,* 749.

Shera, J. (1983). Librarianship and information science. In F. Machlup & U. Mansfield, (Eds.). *The study of information: Interdisciplinary messages*. New York: John Wiley & Sons.

Simpson, G. S. (1962, January). Scientific information centers in the United States. *American Documentation*, 48.

Snow, C. P. (1961). *The two cultures and the scientific revolution.* New York: Cambridge University Press.

Swanson, S. (1995, July 23). Libraries in carpool lane of information superhighway. *Chicago Tribune*, pp. 2C, 1.

Take-off universities (1963, May 24). *Time*, 62.

Thomas, K. (1995, March 19). Libraries opening a new chapter in high technology. *Detroit News*, p. 5C.

U.S. organizations, particularly in the chemical field, use advanced information retrieval on expanding scale. (1961, July 24). *Chemical & Engineering News*, 96.

Welcome to the library of the future. (1996, September 12). *University Library of Case Western Reserve University*, 1.

Western Reserve up-dates information retrieval center (1961, July). *Business Automation*, 40.

White, H. S. (1963a). *The IBM DSD Technical Information Center*, presented at ADI conference, CBI 81, Box 2, Folder 23 (1963). University of Minnesota. Charles Babbage Institute.

White, H. S. (1963b). Mechanized information processing and the librarian. *Canadian Librarian, 19*(2), 68.

Williams, R. V. (1997). The documentation and special libraries movements in the United States, 1910–1960. *Journal of the American Society for Information Science, 48*(9), 781.

Zachary, G. P. (1997). *Endless frontier: Vannevar Bush, engineer of the American century.* New York: Free Press.

Information Retrieval in Science

The Technical Aspects

Examples of Early Nonconventional Technical Information Systems

Madeline M. Henderson

Abstract

Chemists have a long-standing appreciation for the value of recorded information. Many of the early efforts to improve information-processing techniques were centered on chemical information problems. The preeminence of *Chemical Abstracts* as a secondary publication service was well established, but the control and effective use of other information resources were objects of much effort and interest on the part of librarians and information specialists as well as practicing chemists and research scientists, who developed innovative techniques and experimented with the available equipment for handling chemical information.

In the late 1950s the Office of Science Information Service of the National Science Foundation initiated a series of reports (*Nonconventional technical information systems in current use, 1958–1966*) describing some of these innovative information systems. As principal compiler of the first reports, I visited many of the organizations operating such systems. A review of a representative sample of them illustrates the imagination and initiative displayed by the system designers. Thus we can acknowledge the efforts of these pioneering individuals and also recognize their contributions to further system developments.

Introduction

Many of the early efforts in documentation and information retrieval were centered on chemical information problems and possible methods for improved handling of chemical information. The preeminence of *Chemical Abstracts* as a secondary publication service is well known and has been appreciated by chemists over the years. Making other chemical literature resources accessible challenged librarians, information specialists, and practicing or bench chemists and research scientists.

Chemists have always managed to find ways to locate and retrieve data relevant to their interests and current work. Saul Herner (1954) published the findings of his study of ways by which pure and applied scientists obtain information. He found that workers in pure science, including chemists, tended to conduct their own information searches and to ferret out and evaluate the sources they consulted. Improved searching techniques and tools might make these chores easier and more rewarding for them. The applied scientists, on the other hand, seemed to prefer to have their searches done for them; if possible, they wanted references evaluated and summarized.

Herner (1954, p. 235) further suggested that librarians and information officers tend to imagine what the scientist requires: “Too often the scientist goes in one direction in solving his information problems, and the literature specialist goes in quite another direction.” Such results of Herner’s study are largely valid. The technologies have changed but the human factors remain close to those he noted.

The flood of information generated by the boom in scientific and technological research in the 1950s threatened to overwhelm the traditional methods used by researchers to locate desired data. New techniques were being explored, including recording on punched cards of various kinds or recording in new formats, and sorting or searching through the resulting records. The emphasis always included a focus on information content; the new techniques were used to get at the information needed by the users of the systems being established. Such systems were early efforts to process non-numeric data by means usually applied to number crunching—business accounting machines and early computers, for example. The efforts also included examination of the terminology of information resources.

Documenting Nonconventional Systems

During this time, presentations about these early system developments were being made at meetings of the Division of Chemical Literature of the American Chemical Society and of the American Documentation Institute. Papers were being published in such journals as *American Documentation* and *Chemical and Engineering News* and as chapters in such books as *Punched Cards: Their Application to Science and Industry* (Casey & Perry, 1951) and *ACS Advances in Chemistry Series No. 4* (1951). In addition, various reports by industrial organizations and government agencies described efforts in systems' improvements.

These activities, presentations, and publications caused the Office of Scientific Information (OSI; later the Office of Science Information Services) of the National Science Foundation to consider the question of how many of the systems being described were being planned and how many were actually in operation. As the resident chemist on the staff of the Program for Documentation Research in the OSI, I was familiar with the presentations and had visited many of the organizations undertaking the work described in those papers. Sometimes I found working systems, sometimes "drawing-board" plans. We determined in the program that it would be useful to publish a collection of examples of newer methods that were actually being tried. So we launched what became a series of four reports titled *Nonconventional Technical Information Systems in Current Use*. The adjective *nonconventional* was chosen to emphasize that the new or innovative methods were not necessarily mechanized or automated. The introduction to the first issue put it nicely: "systems . . . embodying new principles for the organization of subject matter or employing automatic equipment for storage and search."

Lea Bohnert (1970) did a masterful job of analyzing the series of reports. She noted, for example, that the series progressed from a collection of prose descriptions of systems, through an arrangement of headings defining aspects of the systems, then an organized checklist for system descriptions, to an elaborate questionnaire of nine pages—the System Description Form. I put the first issue together, preparing the prose descriptions for approval by the organizations involved. Herner and Company in Washington, D.C., did the fourth edition under contract. The size also progressed, from 30 systems in the first issue to 178 in the fourth. As a matter of fact, nearly the same number of descriptions appeared in the fourth edition as were reported in the three previous editions and one supplement combined! That growth, plus the indication that the reports were becoming less necessary as a directory of current interest in view of the increasing numbers of references to publications about the individual systems, undoubtedly led the National Science Foundation to stop preparing the series after number 4.

Bohnert also noted that the majority of systems reported small collections of documents, that the systems were located mainly in commercial organizations, and that the dominant subject areas were scientific and technical, with chemical, biological, and medical topics being paramount. She wrote: "It was commercial organizations in the 1950s that pioneered the use of new retrieval methods in small collections . . . because of their own immediate needs and interests" (1970, p. 80).

Those interests included the need to improve access and use of internal company reports and laboratory notebooks, as well as to control data generated by company screening programs and production techniques. In addition, reports from outside organizations, especially from federal government–sponsored work, reprints of note, and patents were included in the files described in these reports. One company maintained a file of research ideas, suggestions submitted by company personnel or contractors relating to new processes or products or new uses for existing products. The file, it was noted, could be used to resolve inventorship questions. Some systems also controlled collections of clippings of interest and information on commercially available products, equipment, and services of interest to the organization. One system, for electronic and electrical engineering products, for example, consisted of folders, one for each manufacturer, into which were placed pieces of literature (brochures, pamphlets, specifications) from that manufacturer. Access to the file was by a coordinate indexing system, using terms descriptive of the characteristics of the equipment covered in that trade literature.

Most of the systems described in our early reports were designed to handle such materials, some more ephemeral in nature than others but all of definite interest and value to the chemists, research scientists, and engineers in the parent organizations. For chemists and chemical organizations the material was more often the internal company documentation, since the published literature of interest was well covered and made available by *Chemical Abstracts*. For organizations not solely or heavily involved in the chemical sciences, which could not count on broad timely coverage of their subject interests such as that provided by *Chemical Abstracts*, the materials in their special collections usually included periodical literature.

Innovative Approaches

Whether working with textual information or numeric data, the new principles for the organization of subject matter adopted in the systems we covered in the reports showed the breadth of innovative approaches offered by systems designers and operators. We had reports on the Zatocoding system, developed by Calvin Mooers' Zator Company. The system was used for internal and external research reports in aeronautical engineering and allied disciplines. This application was a good example of the use of Zatocoding, which worked well in relatively small collections of relatively narrow scope, because the system required analysis and indexing by means of descriptors that were to stand for an idea or concept useful for retrieving information in a particular collection. According to Mooers, descriptors were to be broad in scope, no one subordinate to others, and designed to function independently of each other. The descriptors were carefully derived and defined, and coded as random numbers notched in superimposed fashion on the edge of a Zatocard. To answer questions, a pack of cards was placed in the Zator 800 Selector, rods inserted in a pattern corresponding to the codes for desired descriptors, and the selector vibrated to shake down the cards answering the request. Brenner and Mooers (1958) described this system in more detail in the second edition of the *Punched Cards* book. The system embodied both new principles for the organization of subject matter (descriptors) and automatic equipment for storage and search (the "jiggle box").

A different approach to the organization of subject matter was found in coordinate indexing systems, in which indexing terms were to be coordinated or combined at the time of a search rather than being linked together earlier. One example of coordinate indexing was the Uniterm (a trademark of Mortimer Taube of Documentation, Inc.) system. Unit terms to be used for indexing were to differ from subject headings in two ways: They should be words used by the author of a document, and, as noted above, they should be single terms or single phrases, not modified or compound expressions. The original or standard Uniterm system card contained ten columns in numbered sequence; a document number was posted in the column with the corresponding last digit. The code number for a document was posted on all cards headed by a term contained in the document, resulting in a so-called inverted file. To conduct a search, several cards with appropriate terms were compared, and document numbers in common were identified.

Several of the systems described in the first two reports in our series followed the Uniterm approach. I have mentioned the example of electronic equipment trade literature, which was kept in a Uniterm file. Armour and Company Research Division maintained its file of internal reports and correspondence with a standard Uniterm file. Colgate-Palmolive's Research Department kept the same sort of files, but maintained an auxiliary file of terms useful for broad searches, to which the specific terms used in indexing were related. The example given was the broad term *dental creams*, under which were listed specific trade names to be searched. Another auxiliary file tied specific chemical compounds to more generic compound classes.

Some organizations modified their versions of standard Uniterm systems. For example, Armour used machine posting for its Uniterm file of technical bulletins and clippings on foods, food machinery, chemistry, and chemical technology. A standard IBM keypunch, a sorter, and a 407 tabulator produced duplicate ten-column Uniterm cards. A file of document cards containing all the terms assigned was kept in numerical order, and reference could be made to it rather than to the document. Weinstein and Drozda (1959) published a system description in *American Documentation*. Monsanto Chemical Company's Organic Research Department maintained a Uniterm system for reports, pamphlets, and special publications in the same manner, using the same equipment to produce numerical document lists and ten-column Uniterm postings.

Another variation on the standard Uniterm system was to use internal-punched cards instead of printed cards; searches consisted of comparing term cards for matching holes. This was known as the "peek-a-boo" system; Batten (1951) described an early version in *Punched Cards*. For Armour's patent file, Remington Rand 540-position cards were first used; however, the file was being converted to the larger cards of the National Bureau of Standards (NBS) system. At NBS, Wildhack, Stern, and Smith (1954) had devised a 5-by-8-inch plastic card with provision for 18,000 holes, thus accommodating 18,000 document numbers. At NBS the system was used in the Office of Basic Instrumentation for reports and internal documents on instrumentation.

Another user of the 18,000-hole card was the National Institutes of Health Cancer Chemotherapy National Service Center. The peek-a-boo system was used for indexing and structure searching of the chemical compounds tested in the program for possible activity for cancer treatment. The compound's structure was analyzed, and the structural units to be used for indexing were determined, following the general rules of the center. Searching involved looking for holes common to the cards for desired structural components.

As information specialists and their scientist-users gained experience with the standard Uniterm system and its variations, modifications seemed in order. The developers thus found ways to improve on the improvements. For example, the coordinate index system at Linde Company's Research Laboratory covered internal reports and used IBM punched-card equipment. Fred Whaley, as supervisor of Technical Information Services, found it useful to divide documents into indexable parts and items to minimize "crosstalk," that is, unwanted or inaccurate conjunction of index terms from different parts of the document during retrieval. In addition, each indexing term carried with it a code for the role played by that term in that part of a document. As an example, role 4 was for materials acting as agents of an action and role 42 for a catalyst. The term with its role assigned was called a "structerm."

The punched cards carried the structerm code plus the document number; codes for part and item numbers were superimposed in a field of twenty punch positions reserved for them. Retrieval consisted of pulling appropriate structerm cards and merging them by document number. The resulting decks were compared by peek-a-boo identification of holes at the document part and item level. Whaley (1957) described the system in the published proceedings of a conference sponsored by Western Reserve University.

Whaley's concept of role indicators was adopted for the Uniterm system at E. I. duPont de Nemours Engineering Department. Eugene Wall had initiated the system for internal technical reports, and produced double-dictionary–type index books. Copies were distributed throughout the department for local searching, resulting in a remarkable increase in circulation of reports. System refinements included the use of role indicators (with acknowledgment of Whaley's work) and of links to maintain the relationship among terms. Further developments included converting the index from the dual-dictionary format to punched cards, which could be searched by the IBM 9900 Special Index Analyzer. That equipment, a special-purpose retrieval machine, would be supplanted by the general-purpose IBM 650 computer. Costello (1961) described the system development in *American Documentation*.

Innovative Equipment and Tools

Just as the new principles for the organization of subject matter illustrated innovative approaches, so too did the range of equipment used, from edge-notched cards to computers and everything in between. Edge-notched cards were probably the earliest tools adopted in nonconventional information systems: I have already mentioned Zator cards; McBee Keysort cards were another example. Most of these were 5-by-8-inch cards with holes around the edge, which were notched to represent such items as indexing terms, document numbers, and dates. Inserting a needle into the hole representing a desired item of information allowed for separation of the cards with that hole notched from the rest of the cards. An unusual variation on this theme was employed at the Petroleum Research Corporation in Denver for its file of reports and published articles on the geology of the Rocky Mountains region. Their cards were film transparencies perforated at one end with 207 holes. Code sheets with indexing terms plus pages of the document were reproduced on the card; up to a hundred pages of text could be accommodated. Appropriate holes in the coding area were slotted; selection was by standard edge-notched system techniques. The company produced the card set for purchase by other organizations; it was said that oil companies and the U.S. Geological Survey were using some hundred copies.

Standard machine-sorted punched-card systems were often the next step after the manual systems. At Callery Chemical Company the IBM cards used for its file of data on boron compounds had microfilm inserts. An abstract for the reference coded on the card appeared on the microfilm insert. The abstract was originally typed on a McBee edge-notched card. As that file grew, it was decided to facilitate searching by converting to the IBM system, and the abstracts were photographed for use on the aperture cards.

A unit-card system was used at Union Carbide Plastics Company (formerly Bakelite Company) Development Department for internal technical reports and raw material bulletins. Gilbert Peakes, head of the Development Department Index, noted that a study of the researchers' needs showed that seven major categories would cover all questions. Each indexing term for a given document was punched into a separate card, plus all the information needed for locating the document, including its serial number. The cards were filed by major category. A search consisted of selecting packs of cards corresponding to the codes for desired terms and merging the cards, by machine sorting, into one sequence of serial numbers. Cards with matched serial numbers, found adjacent in the merged set, answered the question. This would seem to be a machine-sorted hybrid unit-term, unit-card system. It was described in a chapter in the second edition of *Punched Cards* (Peakes, 1958).

Ben Weil, as manager of Information Services at the Ethyl Corporation Research Laboratories, developed another use of punched cards. Information on additives used in fuels or lubricants that was found in patents or technical reports was controlled by a Remington Rand punched-card system. A card was punched for each compound mentioned in a reference; a modification of the punched-card code developed by the National Academy of Science's Chemical-Biological Coordination Center was used for the compounds. Each compound card also contained numerical coding for additive functions and types of products. The Remington Rand Bridge (or group selection device) allowed twelve adjacent columns to be sorted simultaneously, so searches could be made for specific compounds or products or for combinations as desired.

At W. R. Grace and Company Research Division, IBM cards were used for a file of company correspondence. Here the sorter was modified by the addition of a ten-column selector device, seemingly similar to the Remington Rand Bridge. In addition, a keyboard control panel was devised to eliminate the need for rewiring the plugboard for each search: an innovative step forward in standard punched-card systems.

The next step up perhaps in the hierarchy of machines used for indexing and searching was the IBM 101 Electronic Statistical Machine. Straightforward applications of such equipment were reported, for example, by the Central Research Department of E. I. duPont de Nemours, Socony Mobil Oil's Research and Development Laboratories, and Union Carbide Chemical's plant in Charleston, West Virginia. At DuPont, chemical compounds in departmental research reports were indexed, characterized by type and number of functional groups, ring structure, configuration, elements present, and so forth. Other subject matter indexed included reaction types and conditions, properties and end uses of the compounds, and miscellaneous information. Each term was coded; both direct and superimposed coding were used on the punched cards. Each card then contained all the information indexed about a single compound. Searches were made both for specific and generic types of information. Edge, Fisher, and Bannister (1957) described the system in *American Documentation.*

At Socony Mobil Oil the punched-card file covered reports, reprints, and patents in petroleum chemistry and technology. The punching scheme provided one or more columns for useful searching categories, with specific headings within the categories assigned to particular holes. Expansion fields were provided for more detail about a particular code. The 101 machine was wired to select cards with the combination of punches corresponding to the combination of index terms. A series of simultaneous searches could be made, with one less term per successive search, to ensure selection of references for broader coverage in case the original search was too specific.

The Union Carbide system was maintained by the Computing Laboratory. Internal technical reports and patents were marked by the technical staff for indexing subjects. Computing lab staff punched IBM cards with corresponding codes, thus controlling the terminology used. Chemical compounds were coded with five-digit serial numbers; more abstract concepts were coded with pattern codes of four pairs of random digits. Serial number codes were punched directly, and pattern codes were superimposed in a field reserved for them. The IBM 101 could select on sixty holes in one pass, directing the cards to any one of twelve pockets, thus making several searches possible simultaneously.

At some organizations the IBM 101 was modified to improve its search capabilities. At Merck Sharp & Dohme Research Laboratories in West Point, Pennsylvania, Claire Schultz managed a coordinate-indexing system in which journal articles, trade literature, and patents in the fields of medicine, pharmacology, and allied sciences were indexed for a mechanized searching system using the 101 machine and random-number superimposed coding. Subject terms and chemical compounds were given random codes and punched into separate ten-column fields on the card. Diseases were coded and punched into another nine-column field. The unique aspect of the system was the special dial board devised by Robert Ford of Merck, which eliminated the need for wiring a plugboard each time a search was to be done. The code numbers for terms defining a search were set (up to four random codes) in the dial board. On one pass the cards were sorted according to the available combinations of codes; answers to more questions than the specific one asked were thus available. This search capability could be said to illustrate the transitional role of such punched-card systems toward later use of computers for information retrieval. Schultz (1958) described the system in a chapter in the second edition of *Punched Cards*.

The Schering Corporation Documentation Center developed an indexing system for literature in pathology and biochemistry that used the IBM 101 machine, fitted out with a similar panel to facilitate searching. The relation of the panel to that at Merck was noted in the system description. Uniterm indexing procedures were followed, and the terms were assigned random eight-digit numbers, punched superimposed in a twelve-column field, or direct codes punched in another five-column field. To accomplish the superimposed punching easily and accurately, duplicate sets of cards punched for each Uniterm and its code numbers were kept; punches for index entries for a given document were copied one after another into a new card (the so-called "lacing operation").

Another system adopting the special dial board was established at Proctor and Gamble's Research and Development Department. Again reference was made to the pioneering work at Merck. The system covered internal technical reports; random number codes for indexing terms were recorded on mark-sensed IBM cards and later punched, superimposed, into cards for searching. There were two fields on the punched cards: One distinguished materials and trade names, and the other distinguished subjects. In searching, the desired code numbers could be set into the special board, with a separate switch set to designate the field to be searched.

Still another system that adopted the Merck special panel with dials was the U.S. Patent Office's experimental system for searching for patents on steroid compounds. At first the office used a special searching machine described as "much like the IBM 101 machine" and labeled ILAS, the Interrelated Logic Accumulating Scanner. It was an example of special equipment that appeared and would later be replaced by general-purpose equipment that could do the same job: In the case of the Patent Office steroid search system, the IBM 101 itself was soon the equipment of choice. The Patent Office search process allowed for setting the dial board to search for as many as seven specific indexing terms; when more were needed, the regular plugboard could be wired. These examples serve to illustrate not only the incentive to improve information systems but also the willingness to share experiences and ideas and the eagerness to learn from one another.

Using Computers for Indexing and Searching

Beyond these applications that used the upgraded IBM 101 Electronic Statistical Machine, our reports contained several examples of the early use of computers for indexing and searching processes. Both general-purpose computers and special-purpose equipment were used in these early innovative systems. We have mentioned the eventual use of the IBM 650 computer in the DuPont Engineering Department system. At DuPont's Textile Fibers Department, internal technical reports were indexed, the index terms coded by randomly assigned alphanumeric codes, the codes stored on magnetic tape, and the coded tape searched by the Bendix G-15D computer. Up to sixteen questions involving sixteen different subject codes could be answered in a single pass of the tape.

At General Electric's Flight Propulsion Division, internal technical reports, memoranda, and government reports were indexed and searched on the IBM 704 computer. The magnetic tape record included key terms plus a brief abstract for each document; a search for specific keywords resulted in a printout of accession numbers of selected documents. That output could be used as input to another run, which printed citations and abstracts of the selected documents.

In 1959 the Cancer Chemotherapy National Service Center contracted Documentation, Inc., to do computer processing of its chemical-biological test data. The system used the IBM 305 RAMAC, or random access computer. It was said that "a program of some elegance" was required for the computations, data reduction, and determination of the status of each piece of screening material. The output of computer runs consisted of printouts of summary data. At the time of the report a program was being written for the IBM 9900 Special Index Analyzer, another piece of special-purpose equipment. At one point in its existence the IBM 9900 was called the COMAC, or continuous multiple access collator.

Both the IBM 305 RAMAC and the IBM 9900 COMAC were used in another contractual system developed and operated at Documentation, Inc., for the Air Force Office of Scientific Research. Descriptions of contracts let by the office were indexed, coded, and punched into cards for collating by the RAMAC and searching by the COMAC. General-purpose computers soon supplanted these special-purpose machines.

Nonconventional Data Files

Another category of information resources covered by our nonconventional systems reports was that of data files: Screening programs for chemicals tested for various applications were tracked and managed in a number of organizations, using different systems and equipment. For example, at the Parke, Davis Central Records Office, internal reports on chemical compounds tested for biological activity were first maintained with McBee Keysort cards and later converted to an IBM punched-card system. Compounds were named according to *Chemical Abstracts* nomenclature rules, and a chemical structure code developed; when the system was on edge-notched cards, the Wiselogle code for structural features was followed. The IBM cards could be copied in answer to structure searches for types of compounds. Harriet Geer, as head of the Central Records Office, suggested that use of the IBM 101 might be the next step in system development.

At Dow Chemical's Central Research Index, data from the agricultural chemicals screening program were handled by standard IBM punched-card techniques. Laboratory notebooks were preprinted so that data could be transcribed easily. Chemical compounds were identified by *Chemical Abstracts* name, as was true at Parke, Davis. Howard Nutting, head of Dow's Central Research Index, noted that chemists thus needed to be familiar with only one naming method to find material in both the open literature and in company documents. At Dow, compounds were also identified by fixed-position codes, and compounds tested were serially numbered. Test organisms, test methods, and biological results of tests were designated by specific code numbers. Two sets of punched cards were prepared, a test set and a chemical set; the two were keyed together by the compound serial number. From the cards, comprehensive lists were prepared by machine in copies sufficient for wide distribution within the company. By the time the system was described in the supplement to the second edition of our reports, there were over one million IBM cards in the main agricultural collection, covering 35,000 chemical compounds, and Dow was conducting research into the use of computers for handling the file.

At Monsanto's Organic Chemicals Division, the IBM 702 computer was already being used for storage and search of data on chemicals produced there and on screening tests. When a Monsanto chemist made an organic compound and assigned a structure and name to it, the information was fed to the computer. A simple coding scheme was designed to convert structural formulas to linear codes, from which the computer could regenerate the structure for display in reports. At the time of the description in our series thousands of screening reports had been prepared and dozens of special compilations printed by the computer, with substantial savings in technical man-hours. Waldo, Gordon, and Porter (1958) described the system in *American Documentation*.

The U.S. Army Biological Warfare Laboratory, then at Fort Detrick, also had files of chemicals, particularly those with herbicidal properties, managed by a Remington Rand machine punched-card system. Three punched-card files, one for empirical formulas, one of chemical groups, and one for visual effects, were used for handling classes of chemicals for structure-function studies. The Remington Rand sorter made such correlations, then final listings were made by a tabulator or, when format was important, on a card-controlled typewriter developed locally. This system had developed over time from a manual system, through edge-notched cards, to the machine-sorted card version. The laboratory expected a computer to be available in 1958 and would study its applicability to the correlation studies.

Lessons Learned

Having reviewed these examples of early nonconventional technical information systems, we can propose some useful lessons to be learned and applied in today's information system design efforts. One of the lessons, I believe, was the recognition of the importance of the information specialist to an organization. A literature chemist, for example, can contribute to research and development efforts of the organization by systematic scanning of the literature pertinent to areas of concern and interest and by preparing bulletins and reports to support researchers and management in planning programs. At Smith, Kline and French Laboratories, research teams were made up of representatives from each of the departments in the Research and Development Division, including the Science Information Department. The information scientist drew upon a broad range of services provided by the Documentation Section of the department. The Information Unit scanned the published literature and prepared weekly abstract bulletins; the Document Services Unit managed the library; and the Document Analysis Unit operated an indexing and storage system for chemical and biological data, using an IBM 101 machine, covering publishede information on the company's products as well as internal reports on these. Henry Longnecker (1956), manager of the Science Information Department, described this approach to effective use of information in the *Journal of Chemical Education*.

Another example of an organization's commitment to improved information use was the provision at DuPont of a group of engineers, supervised by Eugene Wall, who served as documentation consultants for the rest of the company. They assisted in developing various systems, at the time generally similar to the one we described for the engineering department (Wall, 1959). This policy perhaps accounted for the fact that DuPont reported, in our series, the highest number of systems in one organization: seven in the third edition.

We have already spoken of the willingness to share ideas and to benefit from others' experiences. This is as important today as it ever was; we do not operate in a vacuum, and we should keep informed of the broad range of research and development activities going on in information processing. We should also share, through presentations and publications, even in the tentative stages of our thinking and experimenting. Sharing gives us the chance to benefit from the reactions and suggestions of our colleagues, a process well illustrated in the experiences described in our reports. These manual and early machine systems laid the groundwork by defining needs and capabilities that could be embodied in today's systems, which use techniques and technologies only dreamed of by those toiling in the trenches of these nonconventional systems.

References

ACS Advances in Chemistry Series No. 4 (1951). Washington, DC: American Chemical Society.

Batten, W. E. (1951). Specialized files for patent searching. In R. S. Casey & J. W. Perry (Eds.), *Punched cards: Their applications to science and industry*. New York: Reinhold Publishing.

Bohnert, L. M. (1970). Nonconventional technical information systems in current use (NTISCU). In A. Kent & H. Lancour (Eds.), *Encyclopedia of library and information science*. New York: Dekker.

Brenner, C. W., & Mooers, C. N. (1958). A case history of a Zatocoding information retrieval system. In R. S. Casey, J. W. Perry, M. M. Berry, & A. Kent (Eds.), *Punched cards: Their application to science and industry* (2nd ed). New York: Reinhold Publishing.

Casey, R. S. & Perry, J. W. (Eds.) (1951). *Punched cards: Their applications to science and industry.* New York: Reinhold Publishing.

Costello, J. C., Jr. (1961). Storage and retrieval of chemical research and patent information by links and roles in du Pont. *American Documentation, 12*(2), 111–120.

Edge, E. B., Fisher, N. G., & Bannister, L. A. (1957). System for indexing research reports using a punched card machine. *American Documentation, 8*(4), 275–283.

Herner, S. (1954, January). Information gathering habits of workers in pure and applied science. *Industrial and Engineering Chemistry, 46*, 228–236.

Longnecker, H. C. (1956). The role of a science information department in industrial research and development. *Journal of Chemical Education, 33*, 633–635.

Nonconventional technical information systems in current use. (1958–1966). Washington, DC: National Science Foundation. Volume 1, 1958; Volume 2, 1959; supplement to Volume 2, 1960; Volume 3, 1962; and Volume 4, 1966 (prepared under contract by Herner and Co., Washington, DC).

Peakes, G. L. (1958). Report indexing by a unit card system. In R. S. Casey, J. W. Perry, M. M. Berry, & A. Kent (Eds.), *Punched cards: Their applications to science and industry* (2nd ed). New York: Reinhold Publishing.

Schultz, C. K. (1958). An application of random codes for literature searching. In R. S. Casey, J. W. Perry, M. M. Berry, & A. Kent (Eds.), *Punched cards: Their applications to science and industry* (2nd ed). New York: Reinhold Publishing.

Waldo, W. H., Gordon, R. S., & Porter, J. D. (1958). Routine report writing by computer. *American Documentation, 9*(1), 28–31.

Wall, E. (1959, January). Information systems. *Chemical Engineering Progress*, *55*(1), 55–59.

Weinstein, S. J., & Drozda, R. J. (1959). Adaptation of coordinate indexing system to a general literature and patent file: Machine posting. *American Documentation, 10*(4), 122–129.

Whaley, F. R. (1957). A deep index for internal technical reports. In J. H. Shera, A. Kent, & J. W. Perry (Eds.). *Information systems in documentation.* New York: Interscience Publishers.

Wildhack, W. A., Stern, J., & Smith, J. (1954). Documentation in instrumentation. *American Documentation*, *5*(4), 223–227.

Microfilm Technology and Information Systems

Susan A. Cady

Abstract

This paper explores the ways in which the use of microfilm served as a precursor to later computerized information systems in business, education, and science. This accessible and inexpensive photographic technology allowed scientists, scholars, and others to experiment with ideas for information storage and retrieval that were not ultimately realized until the 1990s.

In 1926, George McCarthy, vice president of a New York bank, invented a rotary microfilm camera for copying bank checks automatically. This camera paved the way for the microfilm industry to record large volumes of documents very rapidly. Microfilm was quickly adopted by banks in the 1930s and by other businesses, industries, and government agencies.

Libraries, eager to expand access to resources required by a burgeoning research community, also adopted microfilm. Although academic enthusiasts predicted that microfilm would revolutionize scholarship, the limitations of reading machinery precluded an unmitigated success. Users in academia, business, technology, and science also sought to improve on the ways in which data could be retrieved from microfilm. This paper discusses three specific attempts to overcome output limitations: the struggle for good reading machines; the pursuit of more standardized and sortable formats, i.e., microcards and aperture cards; and finally the use of microfilm as an expendable element in new computerized information systems.

Introduction

Microfilm served as a precursor to and a component of computer-based information systems in business, education, and science. From microfilm's inception as a modern industry in the late 1920s, its proponents extolled its virtues as a solution to many difficult problems in information acquisition, storage, duplication, and retrieval. Sometimes they envisioned it as a complete information system that would revolutionize education, libraries, and the process of scholarly and scientific publication. At a minimum, this highly accessible and inexpensive photographic technology allowed scientists, scholars, and others to "experiment" with ideas for information systems that were only fully realized in the 1990s.

Although microfilm failed to fulfill the loftier visions of its potential, attempts to overcome its limitations as an information medium demonstrate that users wanted more sophisticated information systems. Buckland (1992) and Burke (1994), among others, have documented some of the early attempts at microfilm-based automated information systems. After examining the general social context out of which microfilm's use arose in business, education, and research in the 1930s, this paper discusses three specific attempts to overcome these limitations: the struggle for good reading machines; the pursuit of more standardized and sortable formats, i.e., microcards and aperture cards; and finally the use of microfilm as an expendable element in new computerized information systems.

The term *microfilm* is used here interchangeably with the more comprehensive *microform*, acknowledging that there are many variations on the classic roll microfilm. However, in our context, the term microfilm can conveniently represent all of them except where specifically indicated otherwise. Based on Michael Buckland's exploration of the multiple meanings of information and information systems, microfilm clearly qualifies as an integral part of an information system. In *Information and Information Systems* Buckland (1991) calls for information systems to be broadly defined so that they include a larger universe than the computerized information retrieval systems with which they are sometimes equated. Information systems encompass the selection, storage, retrieval, and output of information. It was the unique ability of microfilm to store large amounts of information inexpensively and for the long term—literally hundreds of years—that attracted users to microfilm in the early twentieth century.

Microfilm in Business and Industry

Microphotography had been invented in 1839 but served primarily as a curiosity for more than seventy-five years. During that period it evolved and improved. By the early twentieth century the push of this advancing photographic technology, readily available even to amateurs in the form of snapshot and home-movie cameras, was augmented by the market pull of large bureaucratic organizations generating ever-increasing quantities of paper documents. The style, quantity, and technology of communication within large business organizations changed dramatically between the mid-nineteenth century and 1920. Managers required frequent, structured written reports and data from their underlings, while executives and supervisors instituted circulars and in-house magazines to communicate downward to subordinates. JoAnne Yates (1989) posits the emergence of a modern system of internal communication as a tool of managerial control. Emerging office technologies, such as the typewriter, duplicating equipment, and vertical filing cabinets, supported this increased activity. Companies expended substantial resources to create, duplicate, file, and store written documents originating inside and outside their businesses; the management of these large volumes of written documents became increasingly difficult.

On the consumer side of the business equation, real income growth, a flourishing advertising industry, such new products as the automobile and radio, and the introduction of installment buying created steady growth, all of which contributed to a dramatic increase in business recordkeeping. Banks promoted checking accounts widely in the early twentieth century, even to industrial workers. But by the late 1920s many banks found the cost of servicing these small accounts so high that they considered instituting service charges (Klebaner, 1990). Among these costs was tracing every check through the banking "transit" system. A 1923 banking textbook noted, "It is essential in banking practice to be able to trace every check handled. It would be ideal to have a photograph of all checks received but manifestly impossible" (Kniffin, 1923, p. 354).

As if in response to this plea, in 1926, George McCarthy, vice president of a New York bank, invented a rotary microfilm camera for copying large volumes of bank checks automatically. Soon thereafter McCarthy granted rights to his invention to the Eastman Kodak Company in return for the presidency of Recordak, a new division created to manufacture and market microfilm cameras. Microfilm was quickly adopted by banks in the 1930s and more gradually by other businesses and industries.

Microfilm in Academia and Science

Interest in microfilm spread to various loosely related nonprofit sectors: the scholarly and scientific communities, research libraries, and government and archival agencies. Libraries, anxious to expand access to resources required by a burgeoning research community, wholeheartedly adopted microfilm. The expansion of library collections has always been a source of pride and competitiveness among institutions of higher education. During the 1920s and early 1930s this competition was intense among major universities, escalating to a sort of intellectual arms race. Faculty and librarians alike knew the stakes. Addressing an audience of alumni, Professor Chauncey Brewster Tinker (1953) of Yale called for action at his institution in these terms:

> If you want your sons and brothers well taught you must have teachers here who are men and learned men; if you are to keep learned men here, you must have a still and quiet place for them to read and think in; but, above all, you must have books for them—not merely a standardized fifty-thousand foot shelf, warranted sufficient for running a university, but a library of millions of volumes, with strange books in it, out-of-the-way books, rare books and expensive books. If we are not willing to compete with the best libraries in this country, it is folly for us to attempt to be one of the great universities, for scholars and teachers, graduate students and at last, undergraduate students will go where the books are. (P. 89)

The output of science and social science research not only added to the pressures on libraries, but also produced a crisis in scholarly communication as delays in the publication of journal articles increased. A diverse company of librarians, archivists, scientists, scholars, philanthropists, and entrepreneurs envisioned microfilm as a partial solution to these problems. Librarians wrote about microfilm, served on professional committees related to microfilm, and attended conferences devoted to microfilm. The tone and volume of this literature reminds one of the early days of computing technology in libraries and even involved some of the same individuals. In 1940, Fred Kilgour, a young assistant at the Harvard University Libraries, wrote in an article for the

Christian Science Monitor that microphotography was "one of the most important developments in the transmission of the printed word since Gutenberg" (Kilgour, 1940, pp. 8–9).

Watson Davis, second director of Science Service, was influential in using microfilm to disseminate scientific and technical journal literature. Founded in 1920 as a nonprofit news syndicate and funded by the Scripps Foundation, Science Service was governed by a board composed of representatives of the American Association for the Advancement of Science, the National Academy of Science, the National Research Council, the E. W. Scripps estate, and the journalism profession. Its mission was to publicize science in the popular press to build support for the continued funding of scientific research that scientists feared would decline with the end of World War I. After he became director in 1933, Davis aggressively pursued his special interest: facilitation of the delivery, publication, and bibliography of scientific literature via microphotography. Inspired by the European-based documentation movement, he adopted as his goal the production and maintenance of "A World Bibliography of Scientific Literature." Exposure to influential librarians, scientists, and inventors in Washington, D.C., provided a personal network and increased his familiarity with both microfilm technology and bibliographic work.

With a $15,000 grant from the American Chemical Society, and the reluctant permission of the Science Service Board, he initiated the Documentation Division of Science Service in 1935 with three projects: operation of the BiblioFilm Service already begun under the auspices of the National Library of Agriculture; development of suitable microfilm cameras and readers to be designed by Navy Lieutenant Rupert Draeger; and microfilm publication of scientific literature by a new entity called the Auxiliary Publication Service.

Because the skeptical Science Service executive committee had specifically limited the subsidy of the new Documentation Division to only fifteen months, Davis acted quickly to ensure its continued existence. He staged a prestigious invitational conference on documentation in January 1937 and orchestrated a call for an organizational meeting to found the American Documentation Institute (ADI). Thus, on 13 March 1937, through his vision and determination, Watson Davis served as midwife to the birth of the American Documentation Institute, the direct predecessor of the American Society for Information Science. He moved the Science Service Documentation Division microfilm activities to ADI. In her book, Irene Farkas-Conn (1989) traces this history in detail.

Other individuals, notably Robert Binkley, Eugene Power, and Vernon Tate contributed to the growth of microfilm in the scholarly world. Professor Binkley, who served as chairman of the Joint Committee (of the Social Science Research Council and the American Council of Learned Societies) on Materials for Research, did an exhaustive study of the efficiency of microfilm as a medium for scholarly publication. His book-length report provides detailed descriptions and actual examples of the film and duplicating technologies of the time; unfortunately his assertions that microfilm could solve the economic problems of scholarly publishing was impaired by his underestimation of the high fixed costs of editorial work that persist in film (and electronic) publication (Binkley, 1936). Eugene Power founded University Microfilm Inc. (UMI) and with the cooperation of large research libraries and foundations began microfilm publication of major research sets, periodicals, and dissertations. *Dissertation Abstracts* began as a free index to UMI's series of dissertations on film. Vernon Tate was deeply involved in almost every aspect of microfilm from the 1930s through the 1960s. First as chief of the National Archives Division of Photographic Resources and Research and then during his twenty-year career as director of libraries at Massachusetts Institute of Technology and the Naval Academy, Tate was instrumental in the founding and growth of the National Microfilm Association. He was heavily involved in several of the early attempts to improve microfilm as an information system by inventing a more acceptable microfilm reader.

The Search for an Acceptable Microfilm Reader

An information system is only as strong as its weakest link, and microfilm was especially weak in retrieving and outputting information to users. Retrieval of discrete information remains difficult on roll film, and, especially in the early days, care was not taken to film materials in a logical order to minimize this problem. So-called "unitized" microfilm (single-sheet film products like microfiche and microcards) with eye-readable headers and automated microfilm systems like Kodak's Miracode in the 1960s attempted to improve retrieval. But the most persistent and difficult problems were with output—the lack of easy-to-use reading machines and printers.

The academic and scientific communities spent decades fostering the design of affordable, comfortable readers without much success. In the 1930s Robert Binkley and others had cast the advancement of science and scholarship as at least partially dependent on widespread adoption of microfilm. Fostering such progress by supporting the development of microfilm technology was thus consistent with the goals of large foundations. The Rockefeller, Ford, and Carnegie foundations funded microfilm laboratories in research libraries and large microfilm publication projects. It seemed logical to enlist the help of the National Research Council, funded by several of the same foundations, to address technical problems relating to microfilm.

The National Research Council established a Committee on Scientific Aids to Learning as one of the technical committees directly under its executive board. James B. Conant, president of Harvard University, served as chairman, and the rest of the membership also was prestigious: Vannevar Bush, then vice president and dean of the School of Engineering at Massachusetts Institute of Technology; Frank B. Jewett, president of AT&T's Bell Laboratories; a well-known New York attorney; two university presidents; one professor, and the chairman of the National Research Council. The committee commissioned a report on the status of microphotography equipment and supplies, the development of specifications for a "student" microfilm reader, and an investigation into eye fatigue. Vernon Tate (1938) was selected to compile the report. Noting that "progress [on readers] has been painfully slow. Little selection is now possible," Tate (1939, p. 44) recommended that the committee assume responsibility for designing a reading device for the individual scholar. The committee did sponsor a competition for a low-cost reader, awarding a contract to the Spencer Lens Company of New York. This device proved inadequate in every way.

In the early days of microfilm some important technical problems with reading machines were solved; for instance, a rotating head was added to facilitate reading materials filmed at both cine and comic orientations. At least one manufacturer, International Filmbook, tried to lessen the high dexterity required for loading roll film into readers by placing the film in individual cassettes. Bankruptcy was Filmbook owner Verneur Pratt's reward for this useful but expensive innovation. Recordak responded to requirements for readers accommodating diverse microfilm formats by offering a variety of machines, including ones specifically designed for newspapers and large engineering drawings on microfilm. World War II interrupted the development cycle; indeed at that time most users were content to have access to any reading device at all, let alone a comfortable and convenient one, since many were requisitioned for the war effort.

Even in the 1960s librarians were still pursuing the low-cost individual portable reader through a project championed by Verner Clapp at the Council on Library Resources. That project floundered for many reasons, including the inability of the library community to standardize microfilm publications sufficiently. Procuring a good reader was a challenge at best; adding the requirement of a "universal reader" for all formats was truly hopeless. Writing to his life-long friend Eugene Power, Vernon Tate (1972) summed up the situation:

> *You* started out with Edwards Brothers in facsimile reprinting of early texts so that they could be made available for people to read. *I* pushed along in microfilm because it was the one way that I could acquire books and manuscripts that I *wanted to read.* One thing is lacking. Microfilm has claimed many adherents who have persuasive arguments for its use in myriad ways, but the truth is that no one wants to read it, really, and it remains at best a substitute form "faut de mieux," and so while so much has been accomplished in some ways, so little has resulted in others.

Microcards: Creating a Standardized Microfilm Information System

Microcards represented another attempt by librarians and scientists to create more robust information systems better designed for retrieval and in a more uniform format. Microcards are opaque cards made of photographic paper, usually 3×5 inches in size, on which page images have been contact printed on both back and front from strips of 16-mm or 35-mm film arranged in rows and columns. Step and repeat cameras that automatically photograph and correctly place sequential pages of text onto the same piece of sheet film were used to film the cards; printing was accomplished by a process similar to consumer photofinishing (Kuipers, 1951). Production of these cards began in the late 1940s and continued into the 1960s when they were largely replaced by microfiche.

The inventor and leading advocate of microcards was Wesleyan University librarian Fremont Rider (1944), whose treatise, *The Scholar and the Future of the Research Library*, dramatized the impending space crisis confronting rapidly expanding research libraries. Rider postulated that the microcard would drastically reduce the four major costs of libraries (original purchase, storage, binding, and cataloging) by combining the catalog card with the reduced text itself contained on the back of the same card. Encouraged by brisk sales of his book, Rider convened a prestigious Microcard Committee composed of appointees of North American library associations and major research libraries. Chaired by Rider, the committee first discussed a centralized publication process and standardization of the microcard itself. Rider emphasized that "the vitally important thing seems to me to be that the library [community] would be able to control microcard standards and so be able to insist upon interfilable uniformity in all microcards" (Rider, 1945a, pp. 162–163). He attempted to put in place an elaborate structure that would ensure this result. A protégé of Melvil Dewey, he was described by one biographer (Parker, 1978) as "a man of singular purpose and enormous drive, not easy to work with and not likely to take note of opposition."

The Microcard Committee debated the feasibility of libraries controlling all aspects of this microformat: selection and physical assembly of items, bibliographical organization (i.e., cataloging), manufacture, and distribution. The group quickly realized that the purchase of capital equipment was beyond the scope of libraries, thus rendering some level of commercial development inevitable. Rider (1945b) hoped that libraries might still uphold these standards by their consumer behavior, since "all they would have to do would be steadfastly to refuse to buy any bastard format, non-standard cards that might be issued by any one." He was adamant that the "library world ought to try to 'direct' the microcarding movement" to avoid dominance by technology developed for other markets, as had so clearly been the case with roll film.

A corporate sponsor emerged swiftly to spearhead and finance the technological developments necessary for microcards, that is, readers and production equipment. Charles Gelatt, chief executive of the Wisconsin-based Northern Engraving and Manufacturing Company, was attracted by a *Time* magazine review of *The Future of the Scholar* and contacted Rider. Gelatt set up the Microcard Corporation in the mid-1940s, and he later formed Micro Library Inc. as a subsidiary to sell the readers and the equipment for producing microcards.

Since opaque materials required reflected rather than transmitted light for reading, reader design was particularly difficult, but several firms developed microcard readers. As of 1950 Rider reported that two readers had been completed: a large standard machine for $195 and a smaller portable machine for $162. By the mid-1960s combination microcard-microfiche readers became available but not combination reader-printers. As mentioned, during the mid-to-late 1950s, the Council on Library Resources financed a long but futile effort to develop a low-cost portable reader for the microcard. With the support of the Microcard Committee, Rider and Microcard Corporation executive Earl Richmond established the Microcard Foundation in Wisconsin in 1948 as a nonprofit organization to coordinate the publishing of materials on microcard. Fremont Rider served as the foundation's chairman from that time until his death in 1962. The foundation supervised the cataloging of microcards to be published to assure bibliographic accuracy acceptable to the library community.

Charles Gelatt had a broader vision for microcard publication than Fremont Rider did for obvious reasons of his commercial interests, and in 1952 his Microcard Corporation contracted to publish Atomic Energy Commission technical reports. Part of the AEC's mission was to disseminate scientific information generated by research into nuclear technology, usually in the form of technical reports. The Atomic Energy Act of 1946 included a directive that "the dissemination of scientific and technical information relating to atomic energy should be permitted and encouraged so as to provide that free exchange of ideas and criticism which is essential to scientific progress" (U.S. Statutes, 1946, p. 755). By 1950 AEC expenditures for research and development had exceeded $120 million, outrun only by Air Force and Navy outlays (Fry, 1953). The AEC Technical Information Service issued declassified wartime reports and unclassified reports of continuing research to the AEC's own laboratories, AEC contractors, and more than forty AEC depository libraries located primarily in major research universities. The Technical Information Center also performed other documentation services, including publication of the indexing and abstracting tool, *Nuclear Science Abstracts*, and the distribution of classified reports to many of the same organizations.

From 1952 to 1964, the AEC distributed twenty million microcards produced by the Microcard Corporation. Microcards were well suited to AEC requirements because individual reports could be more easily retrieved than on roll microfilm. In addition, AEC contractors that had to store classified reports in safes and vaults saved money because of the reduced size. Bernard Fry stated that his estimates were based on conversations with librarians and information officers of agencies; however, at the time of writing Fry himself had been connected with atomic energy research for nine years: initially as an intelligence and security officer for the Manhattan Project and since 1947 as chief librarian of the AEC Technical Information Service.

The Library of Congress Navy Research Section, which soon evolved into its Technical Information Division, issued technical reports on microcards. During the span of microcard production, other types of materials were published as well, including the University of Oregon–sponsored publication of dissertations in health, physical education, and recreation; the famous German chemical reference work *Bielstein*; and even a periodical, *Wildlife Disease,* published exclusively on microcard.

Two features of Rider's original proposal failed to develop as he had recommended in his early publications about the microcard. As noted, one was that the publication of microcards would be undertaken by libraries themselves and not by commercial organizations. The second was that microcards would be filed in the card catalog itself. This impractical approach was abandoned even at his own institution, where the reference librarian announced that "he [Rider] has since decided that ordinarily it is better to store a library's microcards in a separate file, representing them in the catalog with typewritten or L. C. catalog cards. . . . Certainly the removal of cards from a main catalog is never to be encouraged" (Bacon, 1958). Thus, although the rationale for producing cards in a 3×5 size disappeared, the standardization of microcards in this size stood firm, buttressed by the complex support structure established by Rider. The ultimate demise of microcards came not from "bastard formats" of the opaque card but from the revitalization of an older transparent format—sheet microfilm. Unfortunately, many libraries and information centers are still saddled today with almost inaccessible information on microcards, sad remnants of this failed attempt at an improved information system.

Aperture Cards: Sorting Graphical Data

The history of the microfilm format known as the aperture card demonstrates a way in which scientists and engineers developed information systems that would satisfy their needs for storing, sorting, retrieving, and displaying graphic materials, particularly engineering drawings. An aperture card is a device for storing and sorting microfilm copies of paper documents. A hole (aperture) cut in a card is covered and slightly overlapped by a protective glassine sheet adhering to thin strips of pressure sensitive tape around the hole. The glassine sheet is removed when a frame of microfilm is substituted. In a variation without the adhesive the film is suspended from a pocket formed by mounting thin sheets of polyester film on either side of the aperture, leaving one end of one sheet unsealed for the film insertion. Early aperture cards were sometimes mounted on McBee Keysort cards that could be notched on the margins to indicate an index term and then sorted manually with tools resembling knitting needles. In the most widespread application of the aperture card, the microfilm was mounted on an electronic data processing card that could be keypunched and machine sorted.

John F. Langan invented the aperture card while working as chief of the Pictorial Records Division of the Office of Strategic Services (OSS) during World War II. The OSS had appealed to citizens to send the agency any photographs of enemy and occupied Europe that might prove helpful in the war effort. The overwhelming response left the division awash in a sea of millions of bulky, nonstandard, unindexed pictures. In 1943 the agency implemented a system designed by Langan to microfilm the pictures and mount them on aperture cards, thereby reducing these variously sized units to one uniform 35-mm size and facilitating storage, indexing, and retrieval. Near the end of the war Langan filed a patent application for the cards, which was finally granted in June 1950. Failing to interest commercial parties readily in his invention, he sold his rights to two OSS colleagues: Colonel Atherton Richards, former deputy director, and William J. Casey, former chief of intelligence in Europe, and, later, CIA director in the Reagan administration. Richards and Casey incorporated Film 'N File, Inc., contracted with the Dexter Folder Company to manufacture equipment, and arranged distribution through the McBee Company.

A few business applications for aperture cards emerged. For instance, in 1947, the St. Louis Police Department mounted mug shots of criminals on the Film 'N File product. A young Utah engineer designed an effective camera and viewing device for a microfilm aperture card system that was installed in more than fifty real estate title and abstract companies mostly in the western United States. In 1949 Arthur H. Rau of General Electric's engineering division recommended that the cards be used for the storage and retrieval of engineering drawings. Rau estimated that GE alone would purchase millions of them.

In 1951 Film 'N File was renamed Filmsort, Inc., emphasizing the product's ability to manipulate information as well as store it and was acquired by the Dexter Folder Company. Dexter decided to market this product aggressively to the largest user of engineering drawings in the world, the United States government. After an unlikely start, selling the system to the Department of Agriculture for storing and retrieving photographs of meat labels, Filmsort sold heavily into the military—the Air Force, the Army Signal Corps, and the Navy. The services had all used roll microfilm to make archival security copies of engineering drawings, but with the onset of the Korean War they began investigating an integrated system that would dispense with paper drawings altogether in the day-to-day work environment. To illustrate the economy of this approach, the Army Signal Equipment Supply Agency cited its comparison of the same operation of retrieval and refiling of a thousand paper engineering drawings (more than sixteen hours) with the aperture cards (less than five hours) (Davison, n.d.). The ease with which aperture cards could be duplicated facilitated distribution of card sets to multiple sites (satellite plants, technical libraries, vendors, or customers). In addition, multiple card sets filed in different sort orders (by number, by location, by type of machine) enhanced retrievability (Mann, 1976). The Navy contracted with Haloid Corporation for development of modified Copyflo printers to output paper drawings from the microfilm as needed.

Major players in the microfilm and duplicating industries learned that aperture card systems were going to be widely adopted throughout the U.S. military. In the spring of 1954 Remington Rand, Diebold, and Recordak requested and received Filmsort distributorships so that their companies would be ready to supply the cards along with cameras, film, and processing services. In September 1954 Filmsort sponsored a meeting in New York City for the military services, current aperture card users, and suppliers to discuss standards. Participants in the meeting, including the military representatives, agreed that military specifications (standards) for aperture card systems were needed to encourage maximum compatibility between the card systems and supporting equipment. Like Fremont Rider, they understood that the proliferation of formats was potentially a major detriment to economy of operation.

As an outgrowth of this meeting the Department of Defense (DoD) formed its 0009 Committee, and for several years the major players (Navy, Air Force, Signal Corps, Central Intelligence Agency, Western Electric, Recordak, Remington Rand, IBM, Haloid, RCA Victor, Graphic Microfilm, and Filmsort) discussed features like reduction and enlargement ratios, frame sizes, and aperture locations (MacKay, 1966). Even before the final specifications were actually issued in April 1960 (MIL-STD-804, MIL-M-9868, and MIL-O-9878, all 1960 specifications relating to the production of aperture cards, covered sixty-five pages), the committee's work served to rationalize the aperture card and equipment industry by providing de facto standards. The 1961 National Microfilm Association Convention featured multiple sessions on the resulting DoD Defense Engineering Data Micro-reproduction System (EDMS or EDMS-0009) and in 1962 presented its annual award for outstanding achievement in microfilm to the DoD in recognition of this standardization.

In 1959 the Minnesota Mining and Manufacturing Company (3M) acquired Filmsort just as military and industrial use of aperture card systems expanded. The success of aperture card systems in automated information environments would have been seriously threatened if this acquisition had not taken place because, as machine manipulation of cards increased during the early 1960s, the Filmsort adhesive proved inadequate. The adhesive bled beyond the edge of the glassine cover causing the cards to stick together to such an extent that failure rates approaching 30 and 40 percent were common until 3M's adhesives group developed a new substance that eradicated this problem. Although 3M's acquisition of Filmsort was originally motivated by the match between product lines (aperture cards, mounting equipment, readers, and reader-printers), its core competency in adhesive technology cemented the relationship.

Over time many engineering firms, military agencies, contractors, and governmental units incorporated aperture card-based information systems into their operations. Large engineering systems combined the aperture card with the Xerox Copyflo printer. By 1960 numerous large projects were under way: Chrysler's Missile Division distributed more than three million aperture cards with engineering drawings for one of its missile programs; the Army's Redstone Arsenal converted a hundred thousand drawings to aperture cards; military contractor Raytheon Corporation microfilmed more than one million engineering drawings and documents under contract with NATO.

In its ideal application aperture card systems consisted of effective indexing systems, microphotographic technology, electronic card sorters (the workhorses of early data processing), and specialized printers. Users could retrieve individual cards or card sets based on index terms keypunched into the eighty-column card. Thus, the information itself in text or image was more closely linked than in previous systems. In actuality, many users seemed to gain significant benefits simply by maintaining manual filing systems of aperture cards in lieu of the unwieldy engineering drawings. Users may not have had convenient access to card sorters or they may have been discouraged by warnings about damaging cards. Successful card sorting required close tolerances and specialty adhesives—even IBM recommended that sorting be limited to occasional file maintenance. The location of the aperture in the Military D specifications was based on Remington Rand equipment rather than IBM card sorters because the DoD Committee felt this location maximized the space for keypunch data (Carroll, 1960).

Minimizing machine sorting reduced wear and tear on the somewhat fragile film and card systems. In addition, card "scruffing" occurred when the cards slid against each other in sorting equipment. It would be more than thirty years after the promulgation of military specifications for aperture cards and many, many generations hence in computer and telecommunications technology before electronic retrieval of graphic materials like engineering drawings would be achieved. 3M is still active in this market today supplying modern networks of engineering data available to users immediately on the shop floor.

Role Reversal: A Storage Technology Becomes an Output System

A final example of the part microfilm has played in the development of more modern computerized systems is one in which microfilm essentially reverses its role from that of permanent storage mechanism to serve as a temporary inexpensive substitute for paper in an automated environment. The installed base of computers slowly increased in business and industry throughout the late 1950s and then multiplied substantially with the introduction of the IBM 360 machine in the early 1960s. Weak components of computerized information systems became more apparent, and the purveyors of microfilm technology identified two niche technologies that would allow microfilm to serve a role in strengthening computerized systems. In one instance a microfilm innovation served as an input mechanism; in the other instance as an output mechanism. In both cases microfilm built not on its premiere strength as a medium that offered permanent storage but on its usefulness as an inexpensive medium.

The U.S. Census Bureau sponsored the development of microfilm as an input mechanism to a computerized information system. Indeed the Census Bureau's information processing systems beginning with Hollerith's punch card for the 1890 census offer a decennial snapshot of technological progress. In preparation for the 1960 census, Recordak designed specialized high-speed microfilm cameras to feed, flatten, and film coded census booklets at a high rate of speed. Flying spot scanners known as film optical sensing device(s) for input to computers (FOSDIC), developed by the Census Bureau in cooperation with the National Bureau of Standards, then read and recorded codes on the microfilmed documents onto magnetic tape for input into computer systems. The required data were captured and input into computers without extensive handling of paper forms or keypunching. If the spot scanners had to read the codes directly from the paper booklets, dust particles would have prevented the machinery from operating properly: Film was substituted for paper and then discarded.

The system was first used in 1960 and then improved for further use in the 1970 census. The U.S. Weather Bureau modified the FOSDIC system to scan microphotographs generated from three hundred million punch cards of weather information. In 1990 the Census Bureau was still microfilming questionnaires for processing, although FOSDIC itself had been replaced with more sophisticated technologies. Flying spot scanners were used subsequently in true optical character recognition applications, a more advanced version of the technology than detecting marks on microfilmed documents.

In the output niche technology computer-output-microform (COM) served as the medium for the distribution or publication of computer-generated data. The impetus for COM came largely from computer users and potential users frustrated by the slow output of voluminous paper reports. Many businesses believed that computers would not be practical for their applications until printer speeds were substantially improved. Arthur Andersen consultants admitted the gravity of this deficiency when reporting that "not only are the immediate manufacturers of computing equipment engaged in extensive and expensive programs of research designed to produce printers of such fantastically high speeds, but even manufacturers not concerned about the production of computers proper have become active in the development of high-speed printing equipment" (Higgins & Glickauf, 1954).

In COM technology a microfilm camera photographs text and graphic images generated by a computer and displayed on a specialized cathode ray tube (CRT). The CRT was invented in the late nineteenth century and further developed during the 1920s and 1930s for use in television. In the early 1940s the CRT was used by radar developers because it offered a display device with a sufficiently persistent image to compensate for the relatively slow revolution of radar antenna. After the war, defense contractor Stromberg-Carlson unveiled a specialized CRT christened the Charactron to serve as the display device for graphic and textual data generated at high speeds by the Air Force Early Warning System. In the mid-1950s Stromberg-Carlson harnessed its specially shaped tube to the computer at one end, for input of binary data, and to a microfilm camera at the other end for output from the CRT screen to film. Either an online computer link or a magnetic tape supplied the data to the system. A "form slide" was projected onto a mirror between the CRT screen and the camera to provide a template with the traditional row and column divisions and headings expected by report users, without requiring the form itself to be generated on the CRT.

Microfilm-based computer output could be generated at high speed and offered cost advantages over paper where reports or data were widely and frequently distributed, for example, weekly, monthly, or quarterly updates to multiple sites. Until the 1980s to 1990s data users were often not online, so these large outputs were required frequently. In 1961 Convair Aerospace in San Diego first applied COM technology commercially, and in 1966 the Lockheed Technical Information Center in Palo Alto, California, first applied COM in library-related applications. COM products (mostly on fiche) included parts catalogs, price lists, directories, banking data, and managerial reports. The small microfiche reader became ubiquitous in many chain business establishments, including banks. With the advent of machine-readable (computerized) cataloging standards in the mid-1960s, libraries adopted COM technology to produce their catalogs, outputting bibliographic information on roll film that was mounted in specialized readers. With computer-based system development the potential for further improvements of microfilm-based systems was diminished, but the existing microfilm industry found ways to continue its utility. The use of microfilm in these applications also serves as a reminder of how seldom an established technology or an entire information system is completely eclipsed by a new one.

Conclusion

Microfilm, like any information system, has its strengths and weaknesses. One strength is in its ability to preserve information for long periods of time and another is its ability to disseminate information cheaply. The persistence and ubiquity of its parent technology, photography, is an important factor in maintaining this low cost. However, digital photography portends major changes in the technologies used for information storage. Once digital photography penetrates fully into the mass market and new storage mechanisms are developed for digital information that strength may decline. And the World Wide Web offers amazing new facility for instant dissemination of up-to-date information, including the most sophisticated graphic images.

Microfilm's weakness lies in its limited retrieval and output capabilities, both in print and on screen. The repeated, unsuccessful attempts to design comfortable, convenient readers were a testimony to this lack. If today's sophisticated and inexpensive printing technologies had been in existence then, this limitation might have been overcome by converting from screen to paper as is often done today with electronic information. However, the challenge to provide higher levels of retrieval would not be met by even the most sophisticated printer. Microcards and aperture cards were two valiant and very different attempts to improve this aspect of microfilm technology and to provide standardization as well. Although these formats were far from the automated retrieval systems needed to search and rank data quickly and precisely, they did represent an evolution along the way to improved scientific information systems. What the accumulated history of microfilm-based information systems demonstrates is that scientists and engineers, scholars and librarians tried again and again to invent robust information systems because this is what they needed to do their work effectively.

References

Bacon, G. W. (1958). Handling microcards in libraries. *College and Research Libraries*, *11*, 372–373.

Binkley, R. C. (1936). *Manual on methods of reproducing research materials: A survey made for the Joint Committee on Materials for Research of the Social Science Research Council and the American Council of Learned Societies*. Ann Arbor, MI: Edwards Brothers, Inc.

Buckland, M. K. (1992). Emanuel Goldberg, electronic document retrieval, and Vannevar Bush's Memex (Microfilm retrieval technology previous to V. Bush). *Journal of the American Society for Information Science*, *43*, 648–657.

Buckland, M. K. (1991). *Information and information systems*. New York: Praeger.

Burke, C. B. (1994). *Information and secrecy: Vannevar Bush, ultra, and the other Memex*. Metuchen, NJ: Scarecrow Press.

Carroll, H. L. (1960). Use of microfilm aperture cards in IBM equipment. In Vernon Tate (Ed.). *Proceedings of the Ninth Annual Meeting and Convention* (pp. 118–122). Annapolis, MD: National Microfilm Association.

Davison, G. H. (n.d.). *Microtext in the form of microcards and other micro-opaques, transparent microfiches & unitised microfilm or aperture cards as well as reading & enlarging apparatus: Review of progress in 1960*. Rotherham, England: United Steel Companies Ltd.

Farkas-Conn, I. (1989). *From documentation to information science: The beginnings and early development of the American Documentation Institute–American Society for Information Science*. New York: Greenwood Press.

Fry, B. M. (1953). *Library organization and management of technical reports literature*. Washington, DC: Catholic University.

Higgins J. A., & Glickauf, J. S. (1954, March–April). Electronics down to earth. *Harvard Business Review*, *32*, 102–103.

Kilgour, F. G. (1940, September 14). Typography in celluloid. *Christian Science Monitor Weekly Magazine Section*, 8–9.

Klebaner, B. J. (1990). *American commercial banking: A history*. Boston: Twayne Publishers.

Kniffin, W. H. (1923). *Commercial banking: A treatise covering the practical operation of a commercial bank, the theory of money, and banking, and the development of banking in the United States* (1st ed., vol. 1). New York: McGraw-Hill.

Kuipers, J. W. (1951). Microcards and documentation. *American Documentation*, *2*, 89–95.

MacKay, N. (1966). *Hole in the card*: *The story of the microfilm aperture card*. St. Paul: Minnesota Mining & Manufacturing Company.

Mann, J. (1976). *Reducing made easy; elements of microfilm*. Durham, NC: Moore.

Parker, W. W. (1978). *Dictionary of American Library Biography*. Littleton, CO: Libraries Unlimited.

Rider, F. (1944). *The scholar and the future of the research library: A problem and its solution*. New York: Hadham Press.

Rider, F. (1945a). The microcard project. *Journal of Documentation*, *1*, 162–163.

Rider, F. (1945b). The organization of microcard production. *Journal of Documentation*, *1*, 99.

Tate, V. (1938). The present status of equipment and supplies for microphotography: A report prepared for the Committee on Scientific Aids to Learning. *Journal of Documentation*, *1*, 1–62 (Special issue June 1938, no. 3, Part II).

Tate, V. (1972). Vernon Tate to Eugene Power, 12 December 1972. Ann Arbor, MI: University of Michigan, Power Papers, Box 80, Bentley Historical Library.

Tate, V. (1939). Vernon Tate to Keyes Metcalf, 16 December 1939. Ann Arbor, MI: University of Michigan Libraries, National Microfilm Association Records, Box B-2, Eugene B. Power Collection for the Study of Scholarly Communication and Information Transfer.

Tinker, C. B. (1953). The University library. In K. J. Brough, *Scholar's workshop: Evolving conceptions of library service* (p. 89). Urbana: University of Illinois Press. (Reprinted from *Yale Alumni Weekly* XXXIII: 29 February 1924, 650).

U.S. Statutes at Large, *Atomic Energy Act of 1946, 60*, 755.

Yates, J. (1989). *Control through communications: The rise of system in American management*. Baltimore: Johns Hopkins Press.

Mechanical Indexing: A Personal Remembrance

Herbert Ohlman

Abstract

Ideas leading to mechanical and automatic indexing go back 150 years. However, not until World War II, when appropriate tools became available, did these ideas become reality. Early products of these efforts were concordances to full texts, indexes to document collections, and auto-abstracts of documents. I was fortunate to be among those who created one of these systems: permutation indexing. My paper details the development of mechanical indexing and related systems during the 1950s.

Hypertext, which links all types of data, is now the dominant information retrieval technique. It is difficult to remember a time when the character repertory of most computers was limited to all upper-case letters, ten digits, and a few special symbols. However, many researchers realized that these machines were symbol manipulators as well as number crunchers. Specifically, they could be used to process language material, using alphanumeric character sets. Suitable alphanumeric machines, including punched-card (tabulating) machines, punched paper-tape typewriters, and computers became available just before and during World War II. New specialties using these tools were created, notably "information retrieval," "natural-language processing," "speech processing," and "mechanical translation." My paper relates the early history of a subset of information retrieval, "mechanical indexing."

My Work in Mechanical Indexing

My first job in information retrieval (1954–55) was as coordinator of technical information for Carrier Corporation, in Syracuse, New York. At Carrier a marginal punched-card system for technical reports was developed. At Battelle Memorial Institute, in Columbus, Ohio (1955–57), I worked on a team of information retrieval specialists using an effective, but labor-intensive technique called "extracting," developed by Ben-Ami Lipetz. This technique was applied to the documentation and analysis of the literature of titanium for the U.S. Department of Defense (Gibson & Lipetz, 1956). While at Battelle, my first paper was published in *American Documentation* (the original title of the *Journal of the American Society for Information Science*); the subject was the production of marginal punched cards on accounting (tabulating) machines and was based on work done at Carrier Corportion (Ohlman, 1957a).

In late 1957 I joined the System Development Division (SDD) of the RAND Corporation to work on the documentation of one of the largest postwar computing projects: the SAGE (Semi-Automated Ground Environment) air defense system. RAND spun off this division, which became the nonprofit System Development Corporation (SDC). SDD was located first in Lexington, Massachusetts, close to MIT's Lincoln Laboratory, where Whirlwind, possibly the earliest real-time computer, was developed. The librarian at Lincoln Lab's Division 6, Malcolm Ferguson, had installed a peek-a-boo (coordinate) indexing system to retrieve documents. This system used a large card for each index term, in which tiny holes were punched to show that a document used that term. Searches of the collection required superimposing two or more of these term cards over a light source using Boolean logic operations (Figure 1).

Noticing that peek-a-boo cards were sparsely punched, I felt they wasted most of the space (Figure 2). It occurred to me that isolating terms intended by the inherent structure of natural language to be used in context (e.g., words occur in phrases, sentences, or titles) was bound to reduce retrieval relevancy (Ohlman, 1957b).

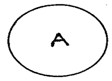
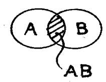
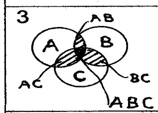
r (no. of coord.) / n (no. of terms)	1	2	3	4	5	6	Σ r
1 	1						1
2 	2	1					3
3 	3	3	1				7
4	4	6	4	1			15
5	5	10	10	5	1		31
6	6	15	20	15	6	1	63
n	${}_nC_1$ =n	${}_nC_2$	${}_nC_3$	${}_nC_4$	${}_nC_5$	${}_nC_6$	$\Sigma\, {}_nC_r = 2^n - 1$

Figure 1. Mathematically possible combinations of n *terms.*

To provide contextual retrieval, I devised a new method based on IBM punched cards and tabulating machines. Significant title words were keypunched, one card to a document title. Tabulating machines were "programmed" by Lewis Hart of SDC (actually, this was done by patch cords on control panels) to punch duplicate cards. In each duplicate card, title words were cycled to the left; the expanded deck was then sorted, and this final deck run through a printer to produce the index. I called the result a "permutation index" because the words went through a cyclic permutation process (Ohlman, 1957c).

The first permutation index was issued by SDC in 1957 as a subject guide to SAGE programming documents (Figure 3). It was based on the titles of 1,800 documents, two-thirds of which were from the Lincoln Laboratory (*Permutation Index No. 1,* 1953–mid-1957). A second edition was issued in 1958, which included 4,000 SDC and Lincoln Laboratory documents. In the introduction I suggested strategies to do quick searches, to broaden a search using "connection-of-ideas" (Whorf, 1956), to get an overview of the corpus of documents, and to find gaps in the corpus (*Permutation Index,* mid-1957–mid-1958).

In 1958 the National Science Foundation, the National Academy of Sciences–National Research Council, and the American Documentation Institute sponsored the first International Conference on Scientific Information. My paper on superimposed coding was accepted for this conference (Ohlman, 1959), and I received preprints of all conference papers.

Here was a perfect way to demonstrate the speed and automation features of permutation indexing to information science and technology colleagues. At SDC, colleagues (Joan Citron and Lewis Hart) and I produced *A Permutation Index to the Preprints of the International Conference on Scientific Information* (Citron, Hart, & Ohlman, 1958). Entries were selected not just from titles; they included author names and affiliations, headings, captions, sentences, and even phrases selected for their significance as thought units. These excerpts provided an average of five permuted entries for every one of the 1,400 preprint pages. Tabulating machines produced the final index but with an improved appearance. Whereas the system used to index internal SDC documents had words truncated to fit into fixed fields, the new index let text flow naturally. The vital alphanumeric character making this possible was the space, which could be used on tabulating machine control panels to determine where to generate additional cards. Also instead of printing the index so that the look-up word appeared at the left margin, it was put in the center to provide context on both sides (Figure 4).

Shortly after the conference, Lewis Hart worked with G. R. Bach to use permutation indexing to analyze verbatim transcripts of psychiatric patients (Hart

& Bach, 1959). However, I found little support for further development of permutation indexing and joined SDC's medical automation project. My thoughts and work on permutation indexing and related matters were described in detail in a paper presented at the twenty-third annual meeting of the American Documentation Institute (Ohlman, 1960).

Other natural-language processing work flourished for another decade at SDC, led by such outstanding researchers as Lauren Doyle, Harold Borko, Robert Simmons, and Carlos Cuadra. This work was described chronologically and in great detail by Doyle (1966) and summarized in books written by authors who were at SDC during the 1957–1966 period (Borko & Bernier, 1978; Doyle, 1975).

Other Mechanical Indexing and Related Systems

At the International Conference on Scientific Information, H. P. Luhn of IBM distributed *Bibliography and Index: Literature on Information Retrieval and Machine Translation* (1958), which contained "titles indexed by Key Words-in-Context system," subsequently known as KWIC (Luhn, 1959). The appearance of the pages of this index were almost identical to the SDC permutation index. Also about 1957 the Rocketdyne Division of North American Aviation developed a system called "rotational indexing" (Carlsen, Garner, & Marshall, 1958, pp. 19–22). However, both IBM and Rocketdyne used digital computer programs rather than tabulating machines to process the keypunched information.

Even before World War II information storage and retrieval systems were developed using similar techniques, but people instead of machines were used to manipulate the information (Yardley, 1931, pp. 255–258; Bernier, 1957; Netherwood, 1958). The war spurred the use of machines to emulate the process (Shera, 1966, p. 79; Veilleux, 1962). Other early efforts were devoted to the production of concordances (Busa, 1951; Ellison, 1957) and to statistical analyses (Heumann & Dale, 1957). However, the earliest reference to permutation techniques was *The Art of Making Catalogs of Libraries* (Crestadoro, 1856; Metcalfe, 1957, pp. 29, 47).

The developers of the Rocketdyne system, a UCLA librarian, and some SDC colleagues and I started a

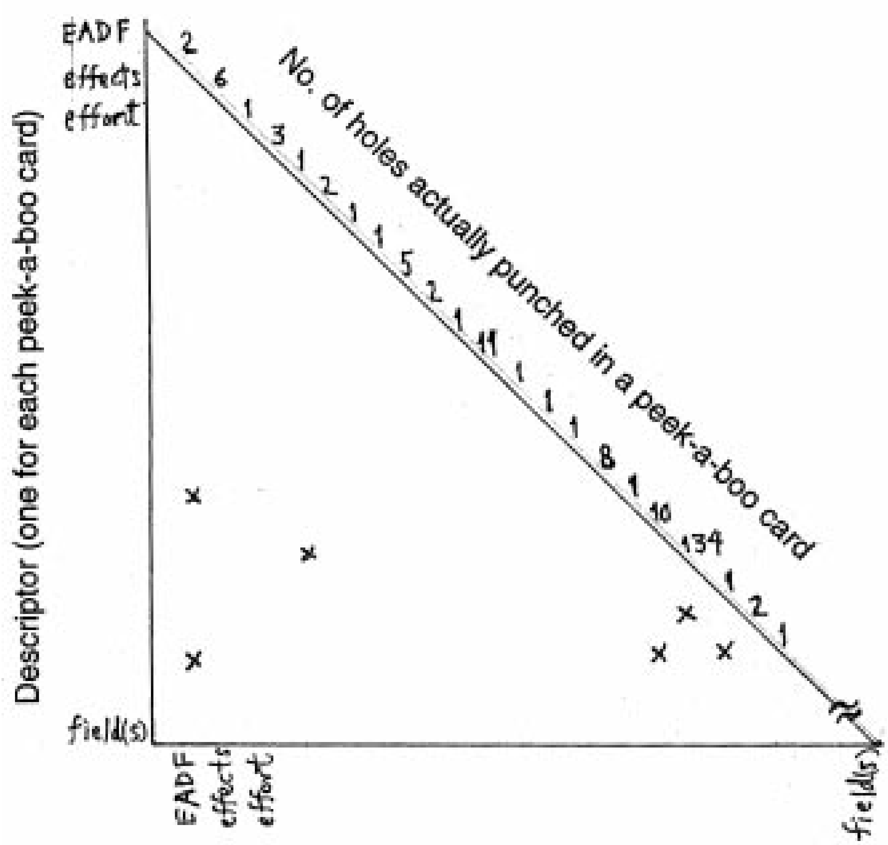


Figure 2. Actual coordinates of 50 terms (r = 2).

Identification Code: *	1	7	6	5	4	3	2	Title starts in column:
5RMRN-062	A-2	COMPILER	SYSTEM				UNIVAC	2
5IBMT0028	ABSTRACT	CODING	SYSTEM	FACS		FLOATING	DECIMAL	3
6ACMJ3129	ACADEMY OF	SCIENCES	USSR		ELECTRONIC	CALCULATIO	MACHINE	4
5ACMJ2095	ACCEPTANCE	TESTS	USED ON	MIDAC			MAINTENANC	2
7MITL5	ACCESSIONS	LIST	MONTHLY				DIVISION 6	2
7RAND9	ACCESSIONS	LIST	WEEKLY		RAND	DOCUMENT	CONTROL	4
7ACMJ4245	ACCOUNT	IDENTIFICA	AUTOMATIC	DATA PROCE				1
3IBMC-210	ACCOUNTING	MACHINE					407	2
2IBMC-035	ACCOUNTING	MACHINES	24	CARD	PUNCH	26	PRINTING	1
3IBMC-019	ACCOUNTING	MACHINES	SORTERS	82	80	75		1
6IBMC-015	ACCOUNTING	MACHINES	CARD	INTERPRETE	550	551	552	1
5IBMT0253	ACCUMULATO	650	MDDPM				INDEXING	2
7ACMJ4005	ADDRESS					INAUGURAL	PRESIDENTI	3
7ACMJ4001	ADDRESS					RETIRING	PRESIDENTI	3
6ACMJ3169	ADDRESS	CALCULATIO					SORTING BY	2
7ACMJ4274	ADDRESS	COMPUTERS			PROGRAMMIN	TECHNIQUES	1-PLUS-1	4
6ACMJ3309	ADVANCED	CODING	TECHNIQUE			USING THE	PACT I	3
5IBMT0149	AIRCRAFT		650	MAGNETIC	DRUM	DATA PROCE	MARQUARDT	6
6ACMJ3299	ALLOCATION	DATA	STORAGE	PACT I			SEMIAUTOMA	2
9IBMC-039	ALPHABETIC	COLLATOR					89	2
4IBMC-021	ALPHABETIC	INTERPRETE					557	2
6ACMJ3175	ALPHAMERIC	INFORMATIO	701			GENERAL	SYSTEM FOR	3
5ACMJ2092	AMPLIFIERS					TESTING OF	OPERATIONA	3
6ACMJ3186	AMPLIFIERS	POTENTIOME		TRANSFER	FUNCTION	SIMULATION	BY MEANS	5
5ACMJ2267	ANALOG	DEVICES				CORRELATIO	COMPUTATIO	3
7ACMJ4012	ANALOG	DIGITAL	COMPUTATIO				SYSTEM FOR	2
5ACMJ2083	ANALOG	INTERPOLAT	AUTOMATIC	CONTROL				1
5ACMJ2028	ANALOG	STUDY OF	ELECTRON	TRAJECTORI				1
6ACMJ3325	ANALOGUE	SOLUTION	PARTIAL	DIFFERENTI	EQUATIONS	HIGHER ORD	DIFFERENCE	3
6ACMJ3085	ANALYSIS					BIBLIOGRAP	NUMERICAL	3
6IBMT+085	ANALYSIS			ITERATIVE	TECHNIQUE	MULTIPLE	CORRELATIO	5
5IBMT0229	ANALYSIS	650	FOURIER	SYNTHESIS	X-RAY	CRYSTAL	STRUCTURE	7
6ACMJ3360	ANALYSIS	DIGITAL	COMPUTING	NETS			ARITHMETIC	2
5IBMT0192	ANTENNA	PATTERN	CALCULATIO					1
7MITL8	APPLICATIO	LOGIC	6D-307				FUNDAMENTA	2
7ACMJ4030	APPROXIMAT	CONTINUOUS	FUNCTION	CLASS OF	FUNCTIONS		CHEBYCHEV	2
6ACMJ3355	APPROXIMAT	CONTOURS					DIGITAL	2
7ACMJ4024	APPROXIMAT	EXPONENTIA	FUNCTIONS				RATIONAL	2
7ACMJ4341	APPROXIMAT	OVERDETERM	SYSTEM OF	LINEAR	EQUATIONS	BEST	LEAST QTH	3
7ACMJ4018	APPROXIMAT	PARABOLIC	EQUATION	STABLE	IMPLICIT	FINITE	DIFFERENCE	5
4IBMT8067	ARITHMETIC	650					COMPLEX	2
4IBMT8060	ARITHMETIC	650				DOUBLE	PRECISION	3
5IBMT0108	ARITHMETIC	650			AUTOMATIC	FLOATING	DECIMAL	4

*Code:
1st numeral: year (in '50-59 decade)
4 letters: journal or organization
4 numerals: journal, vol., and page; or document no.

Figure 3. Fixed-field permutation index.

commercial venture called Permutation Indexing, Inc. The first project was PILOT (Permutation Indexing of the Literature of Technology). The inaugural issue was to be published monthly starting in January 1959. Unfortunately, despite almost one hundred subscriptions, the company was undercapitalized and did not survive long enough to distribute the first issue. Other organizations, however, soon brought out publications using KWIC, notably *Chemical Abstracts* and other chemical publications. Also Eugene Garfield at the Institute of Scientific Information developed Permuterm for the automatic indexing of *Current Contents*.

References

Bernier, C. L. (1957). Correlative indexes: I. Alphabetical correlative indexes. *American Documentation*, *7*(4), 293–298.

Bibliography and index: Literature on information retrieval and machine translation (1958). New York: The Service Bureau Corporation (This bibliography was based on one by Charles Bourne, Stanford Research Institute).

Borko, H., & Bernier, C. L. (1978). *Indexing concepts and methods.* New York: Academic Press.

Busa, R. (1951). *Varia specimina concordantiarum.* Milano: Fratelli Bocca-Editori.

Carlsen, R. D., Garner, W. H., Jr., & Marshall, H. S. (1958). *Information control.* Canoga Park, CA: Rocketdyne, a Division of North American Aviation.

Citron, J., Hart, L., & Ohlman, H. (1958). *A permutation index to the preprints of the International Conference on Scientific Information.* Santa Monica, CA: System Development Corporation. SP-44; Rev. ed., 1959.

Crestadoro, A. (1856). *The art of making catalogues of libraries.* London: The Literary, Scientific & Artistic Reference Office (facsimile reprint OP 14401, Ann Arbor, MI: University Microfilms).

Doyle, L. B. (1966). *The story of language processing and retrieval at SDC.* Santa Monica, CA: System Development Corporation, N-23059.

Doyle, L. B. (1975). *Information retrieval and processing.* Los Angeles: Melville Publishing Company.

Ellison, J. W. (1957). *Nelson's complete concordance to the Revised Standard Version Bible.* New York: Thomas Nelson & Sons.

```
                                               INSTIT                           PAGE  56

 117 RECOMMENDATION FOR AN INTERNATIONAL INSTITUTE FOR SCIENTIFIC INFORMATION *
 7122 ANNUAL BUDGET FOR AN INTERNATIONAL INSTITUTE FOR SCIENTIFIC INFORMATION *
 7119 THE CHARACTER OF THE INTERNATIONAL INSTITUTE /
 TY , LONDON /  7085 D . J . FOSKETT THE INSTITUTE OF EDUCATION , LONDON UNIVERS
 ING CHESTER BEATTY RESEARCH INSTITUTE , INSTITUTE -          3007 DENNIS A . BRUN
 IO /                     1087 GROUP - CASE INSTITUTE OF TECHNOLOGY , CLEVELAND , O
 ERVICE /                              2202 INSTITUTES PHOTO AND MICROFILM COPYING .
 67 B M CROWTHER SCIENCE ABSTRACTS , THE INSTITUTION OF ELECTRICAL -             2
 NED FROM THE LITERATURE AND THE TYPE OF INSTITUTION /   1037 INFORMATION   OBTA
 THEMATICAL PLAN OF THE ESTABLISHMENT OR INSTITUTION /                  1179 THE
 ON -  1176 APPENDIX 1 ORGANIZATIONS AND INSTITUTIONS COOPERATING IN THE COLLECT
                  1039 LITERATURE - AND INSTITUTIONAL AFFILIATION /

          1110 APPENDIX 2 DATA COLLECTION INSTRUCTIONS FOR PROJECT 957 /
                     5172 CONCEPTUAL AND INSTRUMENTAL FORMS /

 /          3075 MAREK CIGANIK CABLES AND INSULATING MATERIALS RESEARCH INSTITUTE

 NTATIVE /          2094 RANDOM SAMPLE TO INSURE THAT PAPERS SELECTED WERE REPRESE

 IDERATIONS IN THE DESIGN -           5003 INTELLECTUAL PROBLEMS AND EQUIPMENT CON

 OF TECHNOLOGISTS FOUND TO BE POSITIVELY INTERCORRELATED /  1238 CHARACTERISTICS

      2179 THE SECOND REASON , IS VESTED INTEREST /
  INTERNATIONAL COMPARISON WOULD BE VERY INTERESTING /                         118

                                    5297 INTERFIX *
                                    5259 INTERFIXING /
                                    5266 INTERFIXING *
                                    5275 INTERFIXING OF SETS /

 ES .                               5201 INTERLINGUAL COMMUNICATION IN THE SCIENC
 GENT /                             5204 INTERLINGUAL COMMUNICATION DAILY MORE UR
                      5107 A SUGGESTED INTERLINGUAL ANALYSIS OF SYNTAX *

                5091 DOES NOT REQUIRE AN INTERMEDIATE CODE BOOK /

 NTERESTING /                       1187 INTERNATIONAL COMPARISON WOULD BE VERY I
 TRACTING .                         2167 INTERNATIONAL COOPERATION IN PHYSICS ABS
  BUILDING , AN APPRAISAL .         2177 INTERNATIONAL COOPERATIVE ABSTRACTING ON
                                    3055 INTERNATIONAL AGENCIES *
 OGRAMS *                           7032 INTERNATIONAL ORGANIZATIONS AND THEIR PR
                                    7061 INTERNATIONAL ORGANIZATIONS /
                                    7093 INTERNATIONAL - COOPERATION *
 NFORMATION .                    7113 AN INTERNATIONAL INSTITUTE FOR SCIENTIFIC I
 POSAL FOR THE ESTABLISHMENT OF A SINGLE INTERNATIONAL JOURNAL /             7 94 PRO
 NS -  7093 G . - A . BOUTRY SECRETARY , INTERNATIONAL COUNCIL OF SCIENTIFIC UNIO
 CTING BOARD , THE STORY OF A VENTURE IN INTERNATIONAL -       7093 THE ICSU ABSTRA
           7119 THE CHARACTER OF THE INTERNATIONAL INSTITUTE /
 NFORMATION *  7122 ANNUAL BUDGET FOR AN INTERNATIONAL INSTITUTE FOR SCIENTIFIC I
 MATION .             7107 CREATION OF AN INTERNATIONAL CENTER OF SCIENTIFIC INFOR
       7005 DIFFERENCES IN NATIONAL AND INTERNATIONAL ARRANGEMENTS FOR FINANCE /
 NFORMATION * 7117 RECOMMENDATION FOR AN INTERNATIONAL INSTITUTE FOR SCIENTIFIC I
  SUPPORT -.             7025 DIFFERENCES IN INTERNATIONAL ARRANGEMENTS FOR FINANCIAL
```

Figure 4. Sample page from "A permutation index to the preprints of the International Conference on Scientific Information."

Gibson, R. W., Jr., & Lipetz, B. (1956). New look in manual methods. *Special Libraries*, *47*, 106–113.

Hart, L. D., & Bach, G. R. (1959). *Natural language indexing by means of data-processing machines; Observation of the growth of perception protocol.* Paper presented at the Western Psychological Association meeting, San Diego, CA. Santa Monica, CA: System Development Corporation, SP-78.

Heumann, K. F., & Dale, E. (1957). Statistical survey of chemical structure. In G. L. Peakes, A. Kent, & J. W. Perry (Eds.), *Progress report in chemical literature retrieval* (pp. 201–214). New York: Interscience Publishers, Inc.

Luhn, H. P. (1959). *Keyword-in-context index for technical literature (KWIC index).* Yorktown Heights, NY: IBM Advanced System Development Division, RC-127.

Metcalfe, J. (1957). *Information indexing and subject cataloging: Alphabetical, classified, co-ordinate, mechanical*. New York: Scarecrow Press.

Netherwood, D. B., (1958). Logical machine design: A selected bibliography. *IRE Transactions on Electronic Computers*, *7*(2), 155–178.

Ohlman, H. (1957a). Low-cost production of marginal punched cards on accounting machines. *American Documentation*, *8*(2), 123–126.

Ohlman, H. (1957b). *Limitations of peek-a-boo and a new retrieval system: Permutation indexing.* Lexington, MA: RAND Corporation System Development Division.

Ohlman, H. (1957c). *Permutation indexing: Multiple-entry listing by electronic accounting machines.* Lexington, MA: RAND Corporation System Development Division.

Ohlman, H. (1959). Subject-word letter frequencies with applications to superimposed coding. *Proceedings of the International Conference on Scientific Information*, *2*, 16–21. Washington, DC: National Academy of Sciences–National Research Council.

Ohlman, H. (1960). *Mechanical indexing: Historical development, techniques and critique*. Paper presented at 23rd annual meeting of the American Documentation Institute, Berkeley, CA.

Permutation index no. 1: A subject guide to SAGE programming documents (1953–mid-1957). Santa Monica, CA: System Development Corporation.

Permutation index: A document finder for FN, SD, SP, TM, 66D, 6D & 6M series. Experimental issue no. 2 (mid-1957–mid-1958). Santa Monica, CA: System Development Corporation, SD-2845.

Proceedings of the International Conference on Scientific Information (1958, November). Washington, DC: National Academy of Sciences–National Research Council.

Shera, J. (1966). *Documentation and the organization of knowledge*. Hamden, CT: Archon Books.

Veilleux, M. (1962). Permuted title word indexing procedures for a man/machine system. In *Machine indexing: Progress and problems* (Papers presented at the Third Institute on Information Storage and Retrieval, February 13-17, 1961, pp. 77–111). Washington, DC: American University Center for Technology and Administration.

Whorf, B. L. (1956). On the connection of ideas. In J. B. Carroll (Ed.). *Language, thought and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA, and New York: Technology Press of MIT and John Wiley & Sons, Inc.

Yardley, H. O. (1931). *The American black chamber*. Indianapolis, IN: Bobbs-Merrill Co.

Science

and

Information

Some National Perspectives

Soviet Scientific and Technical Information System: Its Principles, Development, Accomplishments, and Defects

Ruggero Giliarevskii

Abstract

The Soviet system for scientific and technical information begins with the founding of the Institute of Scientific Information in 1952 at the Soviet Academy of Sciences, now called the All-Russian Institute for Scientific and Technical Information (VINITI). Gradually, VINITI became a center, around which branch information centers developed in industries and in regional capitals that included the central libraries of the Soviet republics and provinces. In 1966 the Supreme Soviet formulated the basic principles for the development of a scientific and technical information system. Many of these fundamental principles never came to fruition, but by the mid-1960s the system counted 2,500 member organs and 11,500 by the mid-1970s.

However, from its very beginning, the system suffered from flaws that increasingly diminished its functional efficiency. The Soviet Union spent ten times less annually on scientific and technical information support activities than the United States, even though the number of specialists working in the field was almost the same. Thus, only about half the Soviet scientific workers had access to a quarter of the world's scientific and technical results—and then only two years after publication. However, since Soviet economics did not stimulate enterprises to master new methods of production, basically the available information was underused. Today, the transition to capitalism requires a totally new system of scientific and technical information. While the previous system encompassed the entire Soviet Union, it now must extend only over Russia.

Introduction

The Soviet Scientific and Technical Information (STI) system (developed in 1952) was destroyed when the U.S.S.R. disintegrated into fifteen independent republics in December 1991. The Soviet Union's national economy was administered by command and constructed on a departmental basis. This method of control was also reflected in the organizational functional structure and practical activity of the STI system.

Transition to market principles of economy began in Russia in the 1990s: Spheres of private property quickly extend; economic methods of management take root; commercialization imperiously meddles in the activity of STI bodies. However, the total transition to information service on a purely market basis would threaten the development of significant spheres of Russia's national economy. The transition requires an essential change in the previous organizational functional structure. A new system is needed, to be created quickly and to make maximum use of structures and staff from the former system.

We must solve this very difficult problem because information and knowledge are essential resources. In this period of transition, scientific development has become a main direction of public manufacture and production. In Russia, with fewer materials and less energy, even more knowledge and skills are required, particularly since the advanced countries of the world have already entered the information era.

Principles

The Soviet STI system was formed in the U.S.S.R. according to the principles stated by V. I. Lenin in 1918–1922, which mainly applied to political speech but also concerned scientific and technical information. These principles embodied the concept of a government monopoly on information activity. In a society entirely controlled by ideology, these principles have often been used as arguments to substantiate decisions made by the Communist party and government agencies.

During the period of its maximum growth the system was based on the following principles (though not all these principles were formulated explicitly):

- Unified government control of scientific and information activity under the U.S.S.R. State Committee for Science and Technology. The development of a specialized (largely departmental) STI system was delegated to government ministries and departments. Within the republics the STI systems were the responsibility of the councils of ministers, while directors of factories and organizations supervised the work of information departments subordinated to them.
- The structure of the STI system was organized like the national economy. Each management level, from government ministries (departments) to local economy units (enterprises and organizations), corresponded to a certain level in the system. Restructuring in the national economy necessarily caused a reorganization of the system.
- Coverage of all types of documentary sources for all fields of science and national economy. The complete coverage was a proclaimed goal, but in practice there was a wide difference in the degree of coverage in individual industries.
- Specialization of STI agencies based on a rational division of functions. The specialization was twofold: centralized, analytical, and synthetic processing of documents by federal and specialized (and partly territorial) STI agencies and decentralized delivery of information to users, accomplished mainly by interdisciplinary and regional STI agencies and information units at enterprises and organizations.
- Uniform construction of the network and organization of activity of STI agencies and special libraries based on standardized reference information collections (federal, disciplinary, regional, and local).
- Unified classification (indexing) of natural and engineering sciences publications by publishers and editors of special journals and information materials kept by STI services.
- Use of modern technologies (computers, office automation, broadcast, motion pictures, and television) to improve the speed and quality of information services provided to scientists, professionals, and industrial innovators.
- Financing virtually all expenses of information services by government budget. (Some self-sufficiency was required, but in practice it boiled down to shifting funds from one budgeted expenditure item to another.)
- International cooperation in scientific and technical information limited because of the ideological and military-industrial confrontation with economically advanced countries.

The forms, methods, and degree of realization of these principles varied in different phases of the system's development.

First Steps

Until the late 1940s the main sources of scientific and engineering information were publications, obtained by scientists and engineers from publishers or libraries. Gradually, special libraries came to be organized into industry-wide networks.

Although information services were set up at factories and design bureaus and disciplinary information centers in some fields operated independently, their links were sporadic and disorganized. As a result there was large-scale duplication in analytic and synthetic processing of information sources and huge gaps in coverage. The coordinating functions of the State Committee for New Technology and its successor, the State Scientific and Technical Committee, were limited mainly to publishing and disseminating new information in industry. Numerous attempts to publish abstract journals initiated since the 1920s never came to much.

After World War II large files of documentation on military technology, especially rocket and radio engineering systems, were brought to Russia from occupied East Germany. The study of these materials gave powerful impetus to active information work in military industries. The demand for special information in various fields of science and industry intensified when a program to develop nuclear weapons, rocket technology, radar, and technical modernization of all military services was instituted.

This was reflected in a decree of the Council of Ministers, issued on 19 July 1952. The decree founded an Institute of Scientific Information of the U.S.S.R. Academy of Sciences (after December 1955, the All-Union Institute of Scientific and Technical Information, abbreviated VINITI) with the mission to publish an abstract journal providing exhaustive coverage of world scientific and technical literature. In 1956 VINITI began also to publish current-awareness publications

(*Express Information*) with abridged translations of significant articles from foreign periodicals. In 1957 it started the series of monographs *Advances in Science and Technology.* By 1960 the *Soviet Abstract Journal* held the first place in the world through the number of sources covered in its abstracts.

The development of information services at industrial enterprises and scientific research institutes continued. Between 1951 and 1955, 230 information units were created (not counting information services in the defense industry); between 1956 and 1960, this number increased by 1,631. The network of specialized information centers continued to evolve.

The transition to industrial management, according to the regional principle of economic boards, was accompanied by the foundation of regional central scientific and technical bureaus in 1957 and central bureaus of technical information for the industry boards, as well as republic information institutes in the Union republics. A swollen network of regional publishing agencies was inefficient. The Council of Ministers, in its decree of 11 May 1962, required centralization of publishing information materials by specialized central institutes. It required mandatory classification of all publications and materials in natural and engineering sciences according to the universal decimal classification (UDC) by publishers and editors of scientific and engineering journals. The use of the UDC in social sciences was rejected for ideological considerations.

Later decrees issued by the Council of Ministers (14 June 1962, 21 May 1964, and 10 September 1964) created the Central Institute of Patent Information, the All-Union Institute of Technical Information Classification and Coding (VNIIKI), and the All-Union Collection of Standards and Technical Specifications (VIFS). The government introduced coordinated acquisition of foreign literature in natural and engineering sciences purchased for hard currency, and information agencies acquired manuscripts of interest to limited groups of specialists (the manuscripts were received for storage and copies were provided on request). Attempts to circulate unpublished research and development documentation through information channels were continued.

Development

The government decree of 29 November 1966, "On the General National System of Scientific and Technical Information," was a major event in this sphere of activities. It regulated the work of ministries and departments and the governments of Union republics and information agencies of different levels in supplying special information to the national economy. It made the State Committee for Science and Technology responsible for the "management of scientific and technical information in the country." It mandated that all publications in natural and engineering sciences be accompanied by source-supplied abstracts and called for the coordination of the activities in information centers and special libraries based on unified reference and information collections. The All-Union Scientific and Technical Information Center (VNTITsentr) created at this time was required to keep a registry of all unclassified R&D projects, both ongoing and completed; file microphotocopies of progress reports and published abstracts; and provide copies of reports on request.

At the government order the State Committee approved standard organization charts for specialized STI systems, which were to "supply all kinds of information services within their respective subject fields to all enterprises and organizations, as well as individual scientists and experts regardless of their departmental affiliation." It also introduced standard organization charts for interdisciplinary and regional scientific and technical information agencies, calling for creation of information centers based on the information bureaus of industry boards that had been abolished in autonomous republics, provinces, and regions.

In 1968 the All-Union Scientific and Technical Research Institute of Interdisciplinary Information (VIMI) was set up to organize interdisciplinary information exchange in the military industry and to transfer scientific and engineering developments from military to civilian sectors of the economy. The Institute of Scientific Information for Social Sciences was also founded at that time. VIMI became responsible for registration of classified and declassified (except for top-secret projects) research work conducted by the defense industry.

These government decrees resulted in a rapid growth of the network of information agencies. The government decree of 19 July 1971 required VNTITsentr to register classified and unclassified research projects conducted in the defense industry (except for the top-secret work). The VINITI Translation Bureau was converted to the All-Union Translation Center. The VINITI Continuing Education Courses became the Institute of Continuing Education for Information Personnel. The Central Statistical Office of the U.S.S.R. was ordered to prepare lists of STI agencies, including special libraries, every five years.

The growth dynamic of the number of Russian and

Table 1. Growth Dynamics of the VINITI *Abstract Journal* from 1953 to 1990

Branches of Science and Technology	Number of Abstracted Publications				
	1953	1960	1970	1980	1990
Automation—radio electronics (1961–)	—	—	64,015	74,360	100,967
Astronomy (1953–)	1,468	12,850	18,015	19,040	29,777
Biology (1954–)	—	119,971	147,699	120,493	255,648
Computer science (1987–)	—	—	—	—	10,158
Geography (1956–)	—	35,781	43,915	45,317	43,474
Geology (1954–)	—	28,342	39,998	38,359	40,546
Geophysics (1957–)	—	16,510	21,547	24,885	24,557
Mining (1960–)	—	16,973	22,911	21,500	24,498
Publishing—polygraphs (1975–)	—	—	—	4,442	4,276
Information science—informatics (1963–)	—	—	4,244	4,762	6,836
Mathematics (1953–)	455	14,640	25,611	35,592	32,220
Machine-building(1956–)	—	135,545	127,374	130,143	144,850
Metallurgy—welding (1956–)	—	30,394	37,096	44,818	47,739
Mechanics (1953–)	1,140	17,065	33,034	34,558	38,077
Management (1970–)	—	—	1,009	3,113	5,775
Environment protection (1975–)	—	—	—	11,749	20,446
Fire protection (1972–)	—	—	—	8,146	6,473
Transportation (1960–)	—	1,388	58,491	69,944	66,543
Physics (1954–)	—	34,450	65,493	83,890	101,544
Chemistry (1953–)	10,042	109,613	237,011	254,166	214,302
Industry economics (1959–)	—	3,168	8,749	10,137	20,593
Electrical power engineering (1955–)	—	83,288	43,708	55,071	66,567
Total	13,105	658,984	1,000,691	1,094,485	1,314,866

foreign publications covered by the *Abstract Journal* of VINITI is illustrated in Table 1. While the growth of world special literature continued unabated, the data show coverage numbers were stalled at a certain level. The loss of the Soviet's leading position in the world in coverage of the literature by its abstract journals meant that Soviet scientists had access to a dwindling portion of the world information flow. It reduced their capacity to trace advances in foreign science and technology and to use new results in their own work.

One important development was the production of prototypes of computerized copying and other equipment by the Electric Modeling Laboratory of VINITI. The first computerized STI systems were put into operation as a result of research conducted at VINITI. The first nationwide automated STI systems were introduced at VINITI and VNTITsentr, and specialized systems were introduced at Electrical Engineering Institute, Instrument Making Institute, Institute of Electronics and Radio Engineering, Institute of Light Industries, and others.

By the mid-1970s a national information network was largely complete. It encompassed all fields of science, industries, and regions and included main types of STI sources. It realized the basic principles outlined above. However, the results were contradictory. Local organizations were buried under an avalanche of standards regulating every aspect of their information activity. All documents concerning the work of the system had to be approved by ministries, departments, and local enterprises and organizations.

The norm-setting documents regulated collection, analytic and synthetic processing, storage, and dissemination of information: There were world science and engineering literature, patent documentation, technical-normative documentation, translations of scientific and technical literature and documents, ongoing and completed R&D projects, doctoral dissertations, patent certificates for products and industrial processes, know-how, industrial products (industrial catalogs), exhibits at national industrial fairs, computer software, educational motion pictures and newsreels, and results of scientific and engineering conferences, congresses, meetings, symposia, and seminars. However, none of these initiatives was brought to fruition, except for coordinating the purchase of foreign literature. Maintaining a registry of performance indicators of information work, based on annual reports of STI services, simply increased the amount of red tape.

Expansion of computerized information processing became the main phase of further STI system development. A network of computerized information centers was created, with remote access to databases produced by national, specialized, and regional information services to accelerate delivery of data to scientists, engineers, and managers. Principles to create integrated information systems were formulated. They called for one-time input (description, indexing, and abstracting) of source documents, conversion of the results to a computer form, multi-aspect data processing, and subsequent multiple use to meet various information needs. Selective dissemination of data, publication of various secondary services, and retrospective document and data searches (including photofactographic retrieval) were included.

The first stage of the state automated STI system included the participation of forty-seven STI agencies: eight national, thirteen specialized, and twenty-six regional organizations. As a result automated services were developed and databases formed for selective dissemination of information and retrospective search, as well as production of secondary publications. VINITI became the main source of databases in the country. In 1989 it published 241 databases in science and engineering, including 42 bibliographic databases with no abstracts (325,000 documents per year), 196 databases with abstracts (957,000 documents annually), and 3 databases for organic compounds and reactions, chemical structures, and biotechnology.

In 1990 retrospective databases published by VINITI covered 10 million documents. The patents database had coverage of 12.5 million documents; regional institutes had coverage of 500,000; and most industrial centers processed 50,000 to 100,000 documents each. However, because of a shortage of high-capacity magnetic disks, VINITI could provide direct access to just 0.1 percent of its cumulative file. Remote access to databases never developed because of the low throughput capacity of communication links in the country.

Copying documents on request remained a "bottleneck" in the work of information centers in the first version of the system. The need for copying equipment at information institutes and centers was at a staggering 70 to 80 percent, resulting in delays of up to four to five months and undermining the value of bibliographic and abstracting information, even if the initial records were found quickly in the database.

In September 1981 the State Committee issued a new edition of its "Standard Procedures for Automated System Development," which in subsequent years was supplemented by numerous additional documents. By mid-1985 there were fifty-five such regulatory circulars in effect. As a result the committee lost its ability to enforce these standards on numerous automated systems. The procedures for coordination gradually became irrelevant.

An interdisciplinary automated STI system evolved separately from the first stage of the state system. Its function was to integrate automated services of the defense industries with the services at VIMI through dedicated communication links into a star-shaped network. The channels were then linked by VIMI with several major research centers and design bureaus in the defense industry. VIMI thus became a powerful information service center. Its equipment and software (in 1992, up to six ES-1066 computers with external memory up to twenty gigabytes) allowed it to simultaneously process up to three hundred requests and service a network of three thousand to five thousand subscribers. This information center satisfied ten to fifteen million requests per year, comparable to the performance of the biggest information centers of the world.

However, the hope of creating an effective system to transfer new developments from the defense industry into the civilian sector was never fulfilled. In fact, the military industry adopted more new ideas from the civilian sector than vice versa, simply because the financial, material, and technical conditions and the infrastructure of defense industry made it better equipped to introduce innovations.

Numerous attempts by the State Committee (in 1978, 1984, and 1987) to draw up a new general government decree mapping out the development of the system failed because the committee lacked new ideas. It could no longer manage the information system according to the old policies. In 1988 it abolished the old practice of annual official approval of the list of information publications (titles of journals, their sizes, and the subject scope of abstracting and analytic reviews).

International Efforts

On 27 February 1969 the governments of Bulgaria, Hungary, the German Democratic Republic, Mongolia, Poland, Romania, U.S.S.R., and Czechoslovakia signed an agreement to create an International Center of Scientific and Technical Information (ICSTI) in Moscow. In 1973 Cuba and in 1979 Vietnam joined this project. ICSTI was treated as an independent international organization, but in actual fact it was a branch of Comecon. In the 1970s it issued handbooks describing the information services of its member countries and listing their publications, as well as presenting a series of reviews of information work in these countries.

The next stage was the gradual formation of an International System of Scientific and Technical Information, which included seven international specialized systems that processed different types of documents and twenty-two international specialized STI systems. Effective interaction of information services of the member countries was facilitated by the center's standardization efforts. As a result several Comecon standards and "Normative Technical Suggestions of ICSTI" were produced. Some work developed standard information technologies to be submitted to national information centers, providing information services to the Comecon administration and its coordinating bodies. This center was, in fact, a component of Comecon administration similar to the function of Soviet industrial centers as units in the management of the national economy. The center's results were determined largely by the preferential treatment it received, such as higher salaries of employees, funds for freelance experts, and a better equipment base. It was a showcase demonstrating the advantages of socialist division of labor. However, some of its activities were truly effective and retained their value for the future.

Remote access to foreign databases through dedicated communication links, specifically with the shared-use computing information center of the Academy of Sciences, was a new stage in the STI system development. The National Center of Automated Information Exchange, created for this purpose at the All-Union Institute of Applied Automated Systems, was linked with computer networks in other countries through remote communication nodes in Austria and Finland. The practical use of the computer network by U.S.S.R. information agencies, however, was hindered by numerous bureaucratic procedures required for network access.

The dissolution of the U.S.S.R. and the formation of the Commonwealth of Independent States meant the disbanding of the State Committee on Science and Technology. Its control of the information systems was ended. Instead, Russia formed the Administration of Scientific and Technical Information and a Committee for Patents and Trademarks. Before 1988 the Association for the Management of Scientific and Technical Information and Knowledge Dissemination controlled a network of interdisciplinary regional information centers in Russia. It was converted in 1988 to the Special Information Association of the Russian Government Planning Committee. Later it became the Russian Association of Information Resources for Scientific and Technological Development (Rosinformresurs), subordinated to the government of the Russian Federation (according to the government decrees of 28 August 1992 and 6 January 1993). The association operates as an integrated information and technological complex responsible for the maintenance and use of regional information resources in Russia.

Accomplishments and Shortfalls

The main accomplishment of the State STI system at the time of economic reform in Russia was the completion of a four-level network of information services, specializing in different types of documents, acquisition of document collections, database generation, and supply of services to various user groups. Bureaucratic barriers were not a flaw of the system itself but rather the inevitable consequence of the command economy and science management. By the same token it was inevitable that the system structure largely replicated the organization of economic management in the U.S.S.R.

National STI Agencies

National agencies were created at different times by government departments participating in the formulation of the national scientific and technological policies. Allocation of funds for the activities, equipment, and the like increased or decreased depending on the importance assigned by the government to various aspects of technological policy. But there was no general underlying rationale behind these changes.

For instance, the hope that basic science would have a key role in the global rivalry between socialist and capitalist systems led, in 1952, to the formation of VINITI with the mandate to provide an exhaustive coverage of the world literature. The commitment to widespread dissemination of advanced know-how to improve productivity was behind the creation of the giant Exhibition of National Economic Achievements in 1959. To intensify the activity of inventors in industry and research, the Patent Information Center was set up in 1962. The attempt to raise Russian technology "to the best world levels" by means of standardization gave rise to the VNIIKI and VIFS institutes. The decision to organize VIMI in 1968 was prompted by the effort to achieve military and technological superiority over the West. Likewise, the plan to switch economic management to the regional level was behind the formation of centers at industry boards and republic information institutes in Union republics in 1957. Each time, when it was discovered that the next "key link" failed to produce immediate dramatic results, the government switched its attention to a new panacea.

As a result the specialization of national STI agencies directing the descending information flow was based on different principles—partly on subjects and partly on types of document. The national STI agencies still remained the main sources and channels of information delivery to scientists and to other professionals. They had just 10 percent of the information industry workforce and were responsible for production of 74 percent of all publications (including 83 percent of abstracts). They satisfied 8 percent of requests for supply of scientific and engineering documents and their copies.

Turf wars among government agencies were responsible for the lack of centralized publishing of information on world literature. During the early years of its existence VINITI abstract journals were superior to their foreign counterparts, but limited resources (especially scarce hard currency funds) by the latter half of the 1970s stabilized coverage at 1.3 million, and the journals increasingly lagged behind their foreign competitors. The coverage of the world literature in the abstract journals published by federal centers responding for medical, agriculture, and construction information never exceeded 20 percent of the world information flow, providing only fragmentary information to subscribers with no clear selection criteria. These three information centers (VINITI, VIFS, VIMI) largely duplicated the VINITI effort by processing the same periodicals. The abstracts supplied by the Institute for Social Information were also selective because of ideological considerations. Expert assessments of documentary information sources received by the state system in 1987 revealed these inadequacies; while patent documentation was covered to an extent close to 100 percent, just 50 percent of scientific and technical publications were processed. For normative and technical documentation the number was 10 percent, with just 8 percent of information on new industry products available. Some 1,500 important foreign periodicals were not received in Russia at all. Conference proceedings, R&D reports, and dissertations were received sporadically. No information agency was responsible for systematically collecting and processing these documents.

All national STI agencies were concentrated in Moscow, where over a thousand general, special, and technical libraries (not counting the libraries in the defense industry) were located, with multiple duplication of book and journal collections. However, this did not guarantee scientists access to even domestic books or periodicals. A large part of library collections were taken out of circulation because of lack of space, and some materials were simply scrapped. Moreover, the central national library (the Lenin Library) made a negligible contribution to the network: It functioned merely as the information agency on problems of culture and arts. Despite these shortcomings the centralized processing of the main types of documents by national STI agencies enabled them to eliminate duplication in the purchasing and analytic-synthetic (meaningful) processing of the literature (especially foreign publications). This reduced the expenses of specialized and regional information systems involved in the formation of collections and provided access to centralized document files to participant organizations.

Another achievement of the national information agencies was formulation of procedures for analytic and synthetic processing of large flows of documents and computerized generation of a wide spectrum of information products and services. Preservation of the former national STI agencies at the federal level and promotion of their activity with elimination of unjustified duplication of effort remains a key challenge facing scientific and technological policy in Russia. Another important objective is connecting national information centers to communication networks (including the defense network, as has been done in the United States) to enable them to realize the full potential in providing access to information for scientists and other professionals.

Central Specialized STI Agencies

The disciplinary principle in the designing of the state system was officially established by the edict of the Council of Ministers on 29 November 1966 and resulted in rapid growth of the number of information centers in industry centers.

Employing some 11 percent of the personnel of the state system, the industrial centers produced some 20 percent of information publications, including 69 percent of reviews and 100 percent of industry catalogs. Their reference information collections amounted to just 2 percent of the total size of the holdings of the system and largely duplicated materials at VINITI, the Republic Scientific and Technical Library (for scientific and engineering literature), the Patent Library (for patent documentation), and so forth. But they were responsible for just 2 percent of document delivery on request. These central industrial centers provided methodological guidance to information units at enterprises and organizations and helped developed a network of information services. They monitored compliance with national and industry-wide norms and methodologies and identified and promoted new methods of information work. The growth in the number of people employed by these centers was an indirect result of regular personnel reductions at ministries and administrative agencies. While the functions of the ministries and agencies remained unchanged, the staffs were reduced, with some of these individuals simply transferred to the staffs of institutes and design bureaus, including information centers.

The basic concept—each system was to provide all kinds of information service to users in their subjects "regardless of their affiliation"—was never put into effect, mainly because industrial information centers were part of the industrial management system. In reality they not only provided preferential treatment to organizations in their industry but also consistently represented the views of their superiors in the annual reports on the main domestic and foreign achievements in science, engineering, and industry. They wanted to put their industry in a better light and sometimes went so far as to distort the actual state of affairs.

The standard charters of STI agencies called for creating in industries "central scientific research institutes for information and technical economic studies." However, while the standard charters called for technical economic studies "based on information materials (publications and other documentary sources accumulated in specialized information collections)," the ministries and other administrative bodies were primarily interested in estimates of the technical and economic indicators of the activity of enterprises in the industry. The effort of such institutes was increasingly concerned with these tasks, and they became more and more dependent on planning and economic departments of ministries.

Initially, the industrial centers were fully financed from the government budget, but in recent years all kinds of imitation self-sufficiency principles were introduced because enterprises and organizations rather than the individual users paid for the information services. Even after the switch to self-sufficiency was completed for scientific research organizations of the industrial ministries, seventy-eight centers covered 60 percent of their expenses from budgets of their ministries and administrative departments.

By the mid-1980s the activities of these centers and other STI agencies were controlled by such a large number of instructions, manuals, procedures, standards, and other norm-setting documents that it became impractical to monitor compliance. The authorities lost their ability to regulate the development of specialized STI systems. In the course of their evolution many centers became offshoots of the bureaucratic apparatus of ministries, rather than scientific research organizations. The uniformity in the structure and functioning of these networks was largely an illusion.

On the other hand, many of these centers accumulated experience in analyzing documentary sources and compiling reviews of the development of their respective industries in Russia and other countries. Several centers created automatic information services with rapid and purpose-oriented operations. The methodological guidance accumulated by them (mainly concerned with managing information services of subordinate organizations) was of value only as long as command economy methods were still in place.

Republic STI Institutes

The first Republic Scientific and Technical Information Institute (RINTI) was formed in 1954. Other such institutes were created in the 1960s when industry management was organized according to the regional principle.

While initially RINTIs were subordinated to republic industry boards and later, by the late l960s, research coordination committees, they were under the control of planning committees of Union republics and became elements of national economy management in their territories. Their major function was to provide information services to government officials and executives, especially supplying analytic materials (reviews and reference data) to support economic and social management. Publishing secondary documents in the national languages of the republics was another important function.

Also suffering from turf rivalry, RINTI's main objective was to support regional interests in the struggle with central agencies. The privileged position of these institutes in the system is indicated by the fact that they constituted 7 percent of expenditures, while employing just 2.6 percent of workers.

Interdisciplinary Regional STI Centers

Since 1957 central technical information bureaus were set up at industry boards during the period when industry was controlled according to a regional principle.

When the national economic management was reorganized and converted to the specialization principle (in 1965), the government initially decided to close down not only the industry boards but also their information bureaus. After further study this decision was reconsidered: In the Ukraine and Kazakhstan thirteen bureaus were specialized by industry, while in Russia twenty-four bureaus were used as interdisciplinary information centers.

In the past few years the network of regional centers (as well as the network of former All-Union STI agencies) has demonstrated that it can function effectively even during an economic crisis. This vitality is explained by several factors. The large size of the former U.S.S.R. (and today's Russia) required (and still requires) collections of special information accessible to this immense territory, including holdings of patents, standards, and catalogs, duplicating the central holdings kept in Moscow. The rapid pace of industrial innovation in these areas confirms the importance of support services in the Far East and eastern and western Siberia provided by regional patent collections. The proximity of these centers to enterprises and organizations enabled them not only to supply new data and documents, but also to create their own information collections and organize exchange of data among the republics of the former U.S.S.R. and various regions of Russia.

Regional literature and document holdings (formed in the republics of the former U.S.S.R. and regions of Russia) are in demand: Specialists received more than 50 percent of all documents and copies from the collections of central agencies through these services (including libraries). Through selective dissemination of information, republic information institutes and other regional agencies provided their subscribers some 60 percent of new data they received from central organs.

Despite these positive accomplishments there have been some significant shortcomings. An inferior technological base and the lack of skilled personnel have had a negative effect on information quality and speed of service. Important problems remain with the network organization and its evolution as a component of the informational infrastructure necessary for technological progress.

The Rosinformresurs Association, which according to the Russian government decrees (28 August 1992 and 6 January 1993) comprises sixty-nine regional centers, is currently forming local information collections for Russian regions to develop this federal network into an integrated resource. This national network provides access to information sources in various parts of Russia, leaving the local governments responsible for development or for other regional information centers focused on local needs.

STI Agencies of Enterprises and Organizations

Government decrees and orders, the standard charters of information agencies, and other norm-setting documents described information departments and bureaus as structural units that perform the following functions:

- Supply specially prepared information to support decision making in the management of research, development, and industry.
- Supply information to professional users for research and development work, process engineering, and industrial operations.
- Monitor information use by departments of an enterprise and provide information on new technological developments and advanced industrial experience to higher-level information agencies.

In their true form these departments and bureaus could be found only at larger information institutes, design bureaus, and factories. Even then they often combined functions of several services: information, patenting of inventions, technical (design) documentation, and standardization. The number of information departments and bureaus grew rapidly after the edict of the U.S.S.R. Council of Ministers on 29 November 1966. As of 1 January 1990 an average information service at an enterprise had 9.3 employees.

This number hides the actual range, which can run as large as a hundred or more employees at large research institutes and design bureaus in the defense industry to as few as a single full-time information officer at a medium-sized enterprise. Just slightly above 10 percent of 46,000 industrial enterprises had information departments of their own. These departments met 78 percent of requests for primary documents from their own holdings. Many have worked out effective methods for supplying information to users, including local automated services built around databases received from federal, specialized, or regional centers. The market reform requires true independence on the part of industrial enterprises and other organizations. Regulation of information work at the local level is counterproductive, even if within government agencies. At the moment local managers should be able to decide whether to retain, disband, or reorganize their information services.

State Information System as a General Concept

Numerous official documents, with no clearly defined legal status, regulated the specialization of information services (processing, accumulating, and delivering of information extracted from documents) in the country and circulation of the documents themselves. The basic principles of the state STI system were correct; the scope of science and technology and industries and regions that it was supposed to cover made the system unprecedented in world history.

However, the mechanism directing the flows of scientific and technical information had serious flaws. Based either on strict sanctions of the command economy or the interests of user enterprises and individuals, neither mechanism was at work in this system. Each mechanism required additional expenditures, but the outlays on scientific and technical information for the past fifteen to twenty years remained at the annual level of 500 to 600 million rubles.

The U.S.S.R. and the United States had a comparable number of people employed in scientific information, but the Soviet Union spent less on this activity by an order of magnitude. As a result just half of scientists had access to one-quarter of the world flow of scientific and technical literature, with the delay of one-and-a-half to two years, while in the United States 90 percent of all publications became accessible virtually immediately after being issued.

Billions invested in automated systems were not used effectively because the command economy, based on arbitrary decisions, essentially did not need, and in fact was hostile to, objective information. Mistrust of automated systems by the national government was evident in the development of the first stage. A full realization of its potential was also prevented by the shortage of high-capacity magnetic media and fast communication channels and frequent malfunctions in the equipment.

Lack of interest in technical innovation and slow introduction of new engineering concepts results from a general shortage of resources and centralized distribution of funds. The dire state of the industry in 1986, which manufactured just 29 percent of its mass-produced engineering items that met world standards, with 14 percent for the machine tools, and 17 percent for instruments, could not be blamed on a shortage of engineering information.

The main causes of this situation were fourfold: 1) systematic underestimation of the importance of basic science as a foundation of technological progress; 2) bureaucratic barriers between scientific institutions subordinated to the narrow interests of their respective financing agencies; 3) monopoly in technology, engineering, and production of equipment with the specifications defined by the manufacturer rather than the customer; and 4) manufacturers' disincentive to embark on intensified research and development ventures. The latter is indirectly evidenced by the fact that 53 percent of invention applications were rejected by experts for lack of novelty (the comparable figure for the United States is 36 percent). The industrial and technological infrastructures in civilian sectors of the economy were backward, and even the best engineering concepts, when realized, turned out to be shoddy products.

Finally, the intensive militarization of the economy held back Russian science and technology, which continued to lag behind world standards. The potential capabilities of the system in its previous form were not used fully; scientists and other professionals (and sometimes even information workers) were often unaware of these capabilities.

The Future

The development of a national information system for the Russian Federation should proceed from a careful effort to preserve existing information resources and a thorough analysis of the capabilities of the units inherited from the U.S.S.R. system and the needs for information service, primarily in socially important nonprofit spheres. Concepts based on the complete commercialization of STI agencies, treating information as no more than a marketable commodity, are shortsighted.

The new national information system of Russia should be developed in the context of the general improvement of science communications, which include processes of representation, transmission, and production of scientific information in society. These processes form the mechanism of the existence and evolution of science. This implies that restructuring will also affect the channels of scientific and technical information not included in the scientific and technological information system, such as publishing books and journals and use of mass media (radio and television). The improvements of all these forms of information transmission should be fully taken into account as they are affected especially by automated services.

References

Arsky, Yu. M., Giliarevskii, R. S., Turov, I. S., & Chernyi, A. I. (1996). The national structures of scientific and technical information: Russia. In *Infosphere: Information structures, systems, and processes in science and society* (pp. 143–188). Moscow: VINITI. In Russian.

Chernyi, A. I., Giliarevskii, R. S., & Korotkevich, L. S. (1993). National system of scientific and technical information of the Russian Federation: A draft development program. *Scientific and Technical Information Processing, 20* (2), 1–38.

Korotkevich, L. S. (1999). *The state system of scientific and technical information in USSR: Results and lessons* (p. 225). Moscow: VINITI. In Russian.

Mikhailov, A. I., Chernyi, A. I., & Gilairevskii R. S. (1984). The ideas of Lenin on the problems, methods, and forms of the scientific information activity. In R. H. Burger, (Trans.), *Scientific communications and informatics* (pp. 247–260). Arlington, VA: Information Resources Press.

Mikhailov, A. I., Chernyi, A. I., & Gilairevskii R. S. (1984). Integral information systems in the state system of scientific/technical information in the USSR. In R. H. Burger, (Trans.), *Scientific communications and informatics* (pp. 349–360). Arlington, VA: Information Resources Press.

Richards, P. S. (1998). The reconfiguration of international information infrastructure assistance since 1991: The Soviet's international system for scientific and technical information. *Bulletin of the American Society for Information Science, 24*(5), 8–10.

The Soviet Overseas Information Empire and the Implications of Its Disintegration

Pamela Spence Richards

Abstract

Immediately after the founding of the Comecon in 1951, the Soviets began to organize meetings of information workers from socialist countries. Soviet experience had great appeal for developing countries because of Soviet success in transforming a largely illiterate population into a seeming industrial powerhouse. America's image was tarnished in the nonwhite world by continuing legal segregation. Much of the Soviet assistance in encouraging centralization and standardization of information practices abroad was channeled through the International Center for Scientific Information in Moscow. Its stated purpose was to develop the International System for Scientific and Technical Information, which Moscow saw as a vehicle for the inexpensive collection, organization, and dissemination of scientific and technical information throughout the socialist world. The system did ultimately create a set of standards for information formats and numerization for all Comecon countries. The U.S.S.R. also sent out its own experts for on-site technical assistance to information centers in, for example, Hanoi and Havana. Probably the most important method of assistance was the free education that the U.S.S.R. offered thousands of students within the Soviet Union.

Introduction

During the decades immediately preceding 1991, the Soviet Union used its political and economic strength to create an international sphere of influence in fields relating to gathering, organizing, and disseminating information. Soviet initiatives to establish and maintain this sphere of influence were strongest within the "brother socialist states," but they reached beyond the membership of the Council for Mutual Economic Assistance (Comecon) to the numerous "nonaligned" nations of Africa, Asia, and Latin America. Funds were expended on such programs as financial assistance for foreign students' graduate study in library and information science, on the organizing of international conferences on information management, and on donations of technical books to the citizens of emerging countries. In terms of dollars spent, the Soviet overseas international assistance program, which peaked in the early 1980s, dwarfed that of the American government, especially because Soviet expansion came at a time of a rapid decline in American overseas aid programs (Childs & McNeil, 1986, p. 208). This paper will describe the ideological background of the Soviets' information offensive and the methods employed by the Soviets both to build up their information hegemony and to defend it against rivals, chief among them the United States. Finally, this paper will discuss the international implications of the empire's sudden dissolution in 1991, which has left a significant portion of the nonindustrialized world with radically diminished information resources.

Ideological Premises

It would be inaccurate to regard the Soviet information empire as simply a twentieth-century descendant of czarist cultural imperialism, which by 1917 had Russified Eurasia from St. Petersburg to the Kamchatka Peninsula in the Northern Pacific. Czarist expansion limited itself to territories contiguous to and eventually annexed by Russia—even far-flung Alaska actually bordered Russia. The Soviets, by contrast, launched cultural offensives in countries as geographically far removed from Russia as Vietnam, Cuba, and Ethiopia. In another major difference between nineteenth-century Russification and Soviet expansion, the bureaucrats of czarist Russia emphasized Russian orthodox Christianity as an integral part of Russian culture, while Soviet officials emphasized instead the uniqueness of the Soviet Union's experience as the world's first socialist (and officially atheist) country, which they were convinced was of value to the impoverished nations emerging from the yoke of superstition, racism, and imperialism in the 1950s, 1960s, and 1970s. This conviction is the unifying thematic thread that runs through the Soviet scholarly literature on international relations in the 1970s and 1980s (Varaksina, 1976; Gorbacheva, 1981), and it is echoed in the words of scholars from the "socialist brother countries," from Cuba (Le Riverend, 1982) to Vietnam (Bui, 1997).

The Ideological Appeal of Soviet Information Systems

The Soviet socialist approach to information gathering, organization, and dissemination had enormous appeal abroad for a variety of reasons beyond its low monetary set-up and maintenance costs for client nations. Soviet experience in the effective manipulation of scientific and technical information media seemed to hold the key for worldwide socialist industrialization and to the modernization of social infrastructures. As the great colonial empires were dismantled in the decades following World War II, scores of hard-currency–poor new nations were inspired by Russia's twentieth-century transition from feudal absolutism to an apparent industrial powerhouse. The success of the Soviets in wiping out the czarist legacy of mass (75 percent) illiteracy was legendary (Raymond, 1979). The launching of *Sputnik* in 1957 produced worldwide admiration for Soviet technical achievement and seemed to confirm the correlation between Soviet-brand socialism, its characteristic centralized technical information services, and elevated technological productivity.

A speech titled "Lenin's Principles of Librarianship and the Libraries of Socialist Cuba," given in 1982 by the first socialist national library director of Cuba, is a typical expression of admiration for the Soviet model. The director of the José Martí National Library described in 1982 how "we [Cubans] try to copy Lenin's ideas of using libraries to further the revolution by widening reading, stimulating scientific and technical development and awakening a thirst for knowledge" (Le Riverend, 1982, p. 6). The speaker described the lack of support for libraries before Castro's victory in 1959 and the intensive library development and centralization that took place under the umbrella of the national library after the revolution. The national librarian cited the charge of the first session of the Cuban Communist Party for the centralized library system to strive to be more important in "marxist-leninist formation," as well as Lenin's own call for the establishment of chains of libraries and efficient interlibrary loan "so that the people can use every book we have." Under Castro, according to the speaker, Cuba is now "living up to that challenge."

Another factor in the appeal of Soviet socialist information policies was their association with the Marxist doctrine of the international brotherhood of the proletariat, regardless of race. It is hard to overestimate the negative international impact of American racism and the damage to the overseas image of America's material success that was done by continued racial segregation in the United States into the 1960s. Through a barrage of publicity given to institutionalized segregation in American libraries, the Soviet press made it easy for the nonwhite populations of emerging countries to associate American information systems and institutions with American racism. In Soviet eyes this association made the United States peculiarly unsuitable for the leadership of a world that even Harry Truman described as "90% colored" (Sherry, 1995, p. 146). A 1948 article in the Soviet library journal *Bibliotekar'* on "Bourgeois Libraries in the Service of Reaction" (Kozlovskii, 1948a, p. 29) pointed out that only 99 of the existing 734 public libraries in the southern states of the United States had services for African-American readers, adding that "in fact the Negro population of the United States in general lacks the most elementary library services." One month later *Bibliotekar'* returned to the theme of racism in libraries, remarking in a report on the opening of United States Information Services (USIS) libraries in Latin America, sponsored by the Department of State, that "the funds spent on these libraries would be more than adequate to open scores of public libraries for American Negroes, but Uncle Sam's love does not extend to them" (Kozlovskii, 1948b, p. 41). In 1955 a *Bibliotekar'* article titled "Racism in Action" described the beating, arrest, and sentencing in Jackson, Mississippi, of a group of young African Americans who had tried to use the Jackson Public Library. The youths were sentenced to thirty days hard labor and a $100 fine—"a characteristic outcome in contemporary America," according to the article's author (Rasizm v deistvii, 1955, p. 60). By contrast, Marxism, aided by the information systems that produced its apparent efficiency, seemed to offer all peoples, regardless of color, the possibility of access to a dignified existence and material sufficiency.

Beginnings of the Soviet Information Offensive in the 1960s

Within a few years after the founding in 1949 by the Soviets of Comecon, the council began to organize conferences where librarians and information-center directors from member Eastern European socialist countries (Albania, Hungary, Bulgaria, Czechoslovakia, East Germany, Poland, Romania and, after 1964, Yugoslavia) could meet with Soviet colleagues. A prime topic of discussion at these conferences was the centralization of information resources so beloved by the Soviets and so attractive to countries with reserves of hard currency inadequate to pay for multiple duplicative information agencies.

A special word is necessary here to explain the importance attached by the Soviets to the standardization and centralization of socialist information systems. By World War II the Soviets had combined mandatory standardization and centralization with a command economy to compensate for lack of resources and trained manpower. Whatever the inefficiencies of such a system, they were more than counterbalanced—in Soviet eyes—by the enhanced control the system offered. It is these two potential contributions—compensation for inadequate resources and enhanced possibilities for political control—that underlie the (continued) fascination of centralized information systems for totalitarian regimes in emerging nations.

Already in the early 1960s the Soviets launched a series of meetings on centralization for research library directors from socialist countries. The proceedings of these meetings in Budapest (1964), Prague (1966), Moscow (1968), Berlin (1970), Sofia (1972), Bucharest (1974), Warsaw (1976), and Pilsen (1979) are a valuable record of the transition of the socialist countries' library and information systems from, in the words of a Hungarian participant, "old, fragmented systems into efficient centralized systems" (Pudov, 1982, p. 48). At a conference offering a retrospective look at two decades of socialist collaboration in centralization on the Soviet model, a Soviet commentator noted that "the process of restructuring the network of public libraries on the principles of centralization and the questions of realizing the Lenin idea of spreading a unified library system in a nation was becoming urgent in all socialist countries." According to the commentator, all the conference participants were convinced that centralization was the most efficient way of raising the national quality of library service (Pudov, 1982, p. 3).

The stress on centralized information services was part of a larger push for efficient and affordable access to current worldwide scientific knowledge that began in the 1950s with the founding in 1952 of the Institute of Technical Information at the Academy of Sciences (which was transformed in 1955 into the All-Union Institute of Scientific and Technical Information—VINITI) (Richards, 1992, p. 273). Soviet willingness to accept the fruits of Western technology despite its bourgeois origins dates back to Lenin, who maintained the old Russian tradition of respect for and reliance on Western research even after the revolution in 1917. When the Bolsheviks came to power, they were—despite their condemnation of Russia's feudal past—careful not to destroy its scientific institutions. They did not want to repeat the mistakes of the French revolutionaries, who abolished the French Academy as a symbol of the *ancien régime* (Vucinich, 1984, p. 93). The Bolshevik leaders were, on the contrary, champions of conventional science. Before the revolution Lenin had written on the theoretical aspects of modern physics, and Trotsky was a star mathematics student. For Lenin science and its derivative technologies were panaceas for Russia's many ills. The old regime had repressed its development and its norms would replace outworn ideologies and superstition (Graham, 1975, p. 19). The fact that these norms had emerged from the bourgeois West did not trouble Lenin, who chastised as "pseudoradicals" those revolutionaries who believed that communism could triumph over capitalism without learning from and working with bourgeois science (Vucinich, 1984, p. 120).

While the immediate post–World War II years were characterized by Soviet xenophobia and a belittling of Western science, the death of Stalin in 1953 and the ascension of N. A. Bulganin to the premiership in 1955 permitted more overt exploitation of Western research. In 1955 Bulganin proclaimed to the Supreme Soviet, "We cannot forget—and we do not have the right to—that technology in capitalist countries does not stand still, but under the influence of the arms race and capitalists' desire for maximum profit, has, in a number of fields, moved ahead" (Barghoorn, 1960, p. 23). This was a public admission of high-level anxiety about Soviet scientific productivity, which, despite the launching of *Sputnik* in 1957, increased through the 1950s. This anxiety culminated in a 1965 report by Nobel laureate P. Kapitsa to the Academy of Sciences claiming that the productivity of Soviet scientists, as measured by the number of publications per individual engaged in research, was only half that of their American counterparts (Kneen, 1985). Bulganin called for more frequent information exchange with foreign scientists, increased purchases of their technical literature, and wider dissemination of foreign science translated into Russian to improve Soviet productivity. This prioritizing of access to world scientific information explains the phenomenal growth of VINITI, whose charges were: 1) abstracting the world's scientific and technical literature; 2) publishing comprehensive abstracting journals; and 3) conducting research for improving scientific information work. By the mid-1970s VINITI employed over 25,000 workers, published more than seventy abstracting journal series, and was annually reviewing and abstracting one million scientific and technical articles from 25,000 journals in sixty-five languages (Mikhailov, Chernyi & Giliarevskii, 1984).

The Soviet Union's international activities in information were accelerated after the establishment in 1963 of the Comecon's Permanent Commission for the Coordination of Scientific and Technical Research, which included a Working Group charged with the responsibility of raising the professional qualifications of information workers in the socialist member countries. (After 1962 the membership was joined by Mongolia, while Albania ceased participating in the Comecon after 196l.) Before 1970 eleven conferences were organized by this Working Group on professionalism, including one in September 1965 on "the training and continuing education of personnel of scientific and technical information centers of the Comecon." The conference proceedings were usually published in the various national East European bibliographic journals. During this time the Working Group also organized exhibits on information technology and published a dictionary of information terminology (Mezhdunarodnyi Tsentr, 1977, p. 49).

While the Soviet Union focused especially on improving the delivery of scientific and technical information from the 1950s onward, it also encouraged the centralization of cultural information as well. Cultural information management in the "socialist brother countries" was coordinated by an agency called Informkul'tura based in the Lenin State Library in Moscow. It organized conferences and circulated, by exchange or subscription, information on the Soviet Union's activities for its cultural minorities. The Soviet Union, with its diverse ethnic and linguistic populations—supposedly united into a peace-loving and patriotic "homo sovieticus"—considered itself an exemplar of nonracist and enlightened cultural politics for the masses. Research on "culturology" pursued in the U.S.S.R.'s numerous Institutes of Culture informed the work done by the Lenin State Library both in setting library policy for all the country's public libraries and in developing models to be encouraged by other socialist countries. Beginning in the 1960s the Lenin State Library's model of centralizing the direction of cultural information policy in the central national libraries was instituted in Budapest, Prague, Bucharest, Warsaw, Sofia, Havana, and Hanoi (Pudov, 1982, p. 105).

An excellent example of how Moscow could mix cultural politics with continuing education in information professionalism was the expensive conference it organized in 1975 specifically for librarians and information center directors from nonaligned countries in Africa, Latin America, and Asia. At this two-part conference, staged consecutively in Moscow and Alma Ata, Kazakhstan, delegates from sixteen countries listened to speeches by ethnic Kazakhs and Uzbekis on the benefits derived by their cultures from Soviet rule and on the importance of librarians being active in the ideological struggle against capitalist imperialism (Varaksina, 1976, p. 82). Delegates from Egypt, Bangladesh, Venezuela, Guinea, Zaire, India, Congo-Brazzaville, Mexico, Yemen, Peru, Senegal, Syria, Somalia, Tunisia, Sri Lanka, and Ethiopia listened to speeches stressing the important ideological role of public libraries in "forming a communist world view." The delegate from Bangladesh stated that all the emerging countries had the same problems with illiteracy that the U.S.S.R. had experienced before 1917 and spoke of possible Soviet aid in providing audiovisual materials to form "libraries for illiterates." The delegate from Senegal said that Senegal wanted to liquidate libraries for the elite and to convert them into "libraries for the masses." The representatives of each country seemed to vie with one another in declaring how much they had learned from the Soviet experience, and the conference terminated in a united declaration by the participants that "the socialist countries have shown us the way and we must follow it to achieve our goals" (Varaksina, 1976, p. 84).

Expansion in Soviet Aid to International Education in the 1970s and 1980s

The extraordinary expansion of the U.S.S.R.'s program underwriting Soviet higher education for foreign students was part of a larger international cultural offensive begun by the Soviets after the global process of decolonialization began in the 1950s. During this process communism had the advantage of its identification with nationalist, anti-imperialistic forces. While the United States, as an ally of both France and Britain in NATO, seemed to be an heir to the old European system, communism appeared to be a liberating force. But there were rivalries even within the communist bloc: The Soviets began to pay closer attention to the emerging post-colonial countries after the 1955 Conference of Asian and African Peoples staged at Bandung in 1955, where they were shocked by the influence of the Communist Chinese. At this time the Soviet student exchange program was still small: In 1953 the U.S.S.R. spent slightly over $1 billion in foreign communication (including international broadcasting, foreign student aid, trade fair participation, and scientific exchanges). After the 1955 Bandung conference the Soviets prioritized the buildup of the oriental faculties of the Moscow and Leningrad state universities. The agency responsible for the cultural offensive in the newly emerging nations was AGIT-PROP, directly under the Central Committee of the Soviet Communist Party. One of AGITPROP's subdivisions was VOKS (the All-Union Society for Cultural Relations with Foreign Countries). VOKS maintained committees of artists and specialists from all fields who acted as advisers in the selection of materials and representatives to be sent abroad. VOKS's policy was 1) to publicize the achievements of Soviet communism so as to demonstrate material progress and 2) to display sympathy for the cultures of the new nations (Bergen, 1962, pp. 121–125).

What had started in the 1950s as a trickle of foreign students arriving to study at Soviet universities and institutes had become a flood by the late 1970s. An increasing proportion of the total foreign student population in the U.S.S.R. was from Africa, Asia, and Latin America rather than from the European socialist countries. The number of foreign students in the U.S.S.R. from Latin American and the Caribbean more than doubled between 1979 and 1985—from 2,900 to 7,600. Cuba, which by 1980 was receiving $10 million a day in Soviet assistance, was a major supplier of foreign students, as was Nicaragua, where annual assistance from the U.S.S.R. had risen from $6 million in 1980 to $580 million in 1986. In 1985 more than 2,500 Nicaraguan students went to the Soviet Union to study. In that year the largest number of foreign students in the U.S.S.R. was from Bolivia, Colombia, and Costa Rica (U.S. Department of State, 1987, pp. 66–68).

Professional training for foreign students expanded accordingly: In-depth training in information science for foreign students was offered in the Soviet Union after 1963 in months-long continuing education courses set up at VINITI in Moscow (Richards, 1992, p. 275). Support for international information training was stepped up after the founding in 1969 of the International Center for Scientific and Technical Information in Moscow. The International Center was a Comecon institution with a mandate to develop and maintain an international system for scientific and technical information in order to standardize and centralize the information systems of all Comecon countries. A formal Institute for the Raising of the Qualifications of Information Workers (IPKIR) was founded in 1971 and located at VINITI. On the basis of bilateral agreements with various socialist countries (but largely funded by Moscow), IPKIR educated, between 1972 and 1976 alone, 853 students from Bulgaria, Hungary, Germany, Mongolia, Poland, Czechoslovakia, Romania, and Yugoslavia. In December 1975 the International Center organized a large conference for teachers of the theory and practice of information systems, in which delegates from Bulgaria, Hungary, East Germany, Cuba, Poland, and Czechoslovakia participated. The following year a resolution by the Plenipotentiary Committee of the members of the International Center approved an ambitious ten-year program (up to 1985) of internships at IPKIR for higher-education teachers, for the publication of teaching manuals based on symposium proceedings, and for graduate study and assistance points for people teaching about technical information systems. In 1977 an academic department of international systems and technical information was organized at IPKIR. In collaboration with the International Center, IPKIR was to serve for the remaining years of the Soviet Union as a central point for 1) the recruitment of information science trainees for the socialist countries; 2) the pooling of training materials and methods; 3) research on training; 4) lectures by leading specialists; and 5) consultation on training personnel for different national systems of scientific and technical information (Mezhdunarodnyi Tsentr, 1977, p. 50). Because of chronic shortages of resources, however, the continuing education program provided to foreign information professionals by the Soviet Union was not always as elaborate or advanced as it appeared on paper.

Another important site used by the Soviets for subsidizing higher education in library and information science was the Krupskaia Institute for Culture in Leningrad. Named after Lenin's wife, Nadezhda Krupskaia, the Institute in Leningrad annually hosted an average of a hundred foreign students between 1978 and 1985 in its five-year diploma program (Moskalenko, personal interview, 1997). In addition, between 1974 and 1991, the Krupskaia Institute awarded the doctorate (*kandidat*) in librarianship to twelve Vietnamese, two Sudanese, two Cubans, two Syrians, two Afghanis, and individual librarians from Cambodia, Laos, Guinea, Kenya, and Iraq (Dissertations, 1997). (Some foreign students were also educated in the library and information science faculty of the State Institute of Culture in Kiev, but the Moscow Institute of Culture could not be an international training site because of its coincidental closeness to a restricted military zone [Giliarevskii, personal interview, 1997]).

Most of the foreign students who received free Soviet educations were from families of officials in those countries that had bilateral agreements with the U.S.S.R., whereby the U.S.S.R. would give selected applicants a free higher education. The applicants' names were sent to an international section at the Ministry of Higher Education, which made acceptance decisions. New students were sent for ten months to preparatory faculties (*Podfakul'tety*) all over the Soviet Union, which specialized in Russian instruction and adaptation lessons for specific student populations. A *Podfak* in the capital of the republic of Moldova, for example, specialized in the preparation of French-speaking Africans, and the Patrice Lamumba "University of Friendship of the Peoples" in Moscow primarily addressed the needs of Asian, Latin American, and African students. While students were at their *Podfaky*, decisions were made on where to send them for their higher education. The Ministry of Culture then informed the various institutions which foreign students they would be hosting. In the different cities all over Russia where the students were dispersed for their studies, local organizations played an important role in their reception, organizing cultural tours, parties, camping expeditions, and trips. Foreign students at the Krupskaia Institute, for example, were taken on extended boat cruises through Russia's riverways every summer.

It is impossible to overestimate the importance of the Russian experiences of these young people, many of whom spent almost six years in the U.S.S.R. as the privileged guests of the Soviet people. By the time of the dissolution of the Soviet Union in 1991 the Krupskaia Institute alone returned to their home countries nearly two thousand graduate librarians fluent in Russian, versed in Russian culture and geography, and convinced of the advantages of centrally controlled information services on the Soviet model. A number of these graduates rose to positions of importance in their home countries and subsequently influenced the development of local information infrastructures. A Vietnamese graduate of the Krupskaia Institute, for example, is currently serving as director of the National Library in Hanoi (Varganova, personal interview, 1998), and the current director of Vietnam's Center for Scientific and Technical Information studied in Moscow at VINITI (Giliarevskii, personal interview, 1997).

Information Systems as an Ideological Defense against Capitalism

While Soviet-subsidized education in library and information science unquestionably raised professional standards in many of the participants' countries, it also served the purpose of politicizing information work by training librarians and technical information workers to act as "active agents in the class struggle." The importance of this ideological training of information workers for countering the threat of capitalist influence was stressed in a 1981 Soviet report on international socialist collaboration on bibliographic control. The report's author noted that the underlying purpose of all the collaborative activities of the past two decades was "the development of a common socialist culture," which would strengthen the various "brother socialist countries" in three ways: 1) by helping to build a stronger scientific and technical base; 2) by assisting in the development of a "proper orientation" to encroaching Western social ideas; and 3) by "arming the brother socialist countries in their struggle with bourgeois, reformist and revisionist ideologies." The author explained the latter as meaning that librarians in socialist countries needed to evaluate the information streaming in from Western sources "with class consciousness and a partisan approach" (Gorbacheva, 1981, p. 6).

Socialist information workers also needed to be warned of the ideological dangers lurking in Western information technology. A recurring theme of the Comecon library professional conferences of the 1970s and 1980s was the need to counter the overseas influence of MARC (machine readable cataloging), which was expanding its original function of making Library of Congress cataloging machine readable by other American libraries and was becoming an international system for the exchange of bibliographic information in machine-readable form. The Soviets claimed that this enabled the United States to exercise ideological influence on the information activities of participating countries (Gorbacheva, 1981, p. 7).

The Soviet International Information System at High Tide: The MSTNI

The principal task of the International Center for Scientific and Technical Information in Moscow was the establishment and maintenance of a socialist international scientific information network [Mezhdunarodnaia Sistema Nauchno-tekhnicheskoi Informatsii, or MSNTI]. The MSNTI was developed in line with the United Nations Technical Information System (NATIS), created by UNESCO in the early 1970s. NATIS proposed the development of coordinated national scientific and technical information systems that would ultimately become the basis of a global standardized information network, UNISIST. NATIS was based on the principle that the best information on printed materials could be supplied by the countries in which they were produced. UNISIST was conceived specifically to stimulate the creation of national bibliographies for countries without them.

The Soviets intended their own international system to demonstrate superior Soviet experience in information centralization, as well as international Soviet-led collaboration in information science. Furthermore, MSNTI would compensate for the inability of hard-currency–poor socialist countries to pay for multiple copies of expensive Western journals. Ideologically, the MSNTI was justified as a means of supporting the struggle of the masses for peace and disarmament, another of the political themes resonating through the Soviet information literature of the 1970s and 1980s (Gorbacheva, 198l, p. 5; Tvardovskaia, 1984, p. 68). Ultimately, because of the chronic shortage of material resources, the MSNTI never worked in reality as efficiently as it appeared to on paper; on the other hand, it was far from being a sham operation and certainly raised standards in the Soviet client countries.

In the early 1980s the Soviets stepped up their campaign to equate strong socialist information systems with the defense of socialist ideology. At a conference in 1983 at the Lenin State Library on "Librarianship in the Capitalist Countries and the Current Ideological Struggle," participants were reminded of the resolution of the 1983 and 1984 party congresses that "capitalist library theory and practice be relentlessly criticized and that socialist librarians become more active in the formulation of public opinion." They were told that the "spread of American library services all over the world facilitates the infiltration of American ideology. Libraries that use such services will inevitably fall under American control." Electronic databases and the MARC system were cited as the United States' two newest ideological weapons. American "objectivity" was, according to one speaker, simply a pose; American librarianship had always served bourgeois capitalist interests (Tvardovskaia, 1984, p. 78).

Soviet Book Distribution Programs

An important element in the establishment of Soviet influence abroad in many fields was the U.S.S.R.'s support of a massive international book-publishing and distribution program to support its international information offensive. In 1982 alone the Soviet Union produced 74.5 million books in fifty-six non-Soviet languages, a large proportion of these being in scientific and technical fields. That year they published 24.3 million English-language books—more than in any other language the Soviets published. By 1986 one out of every four books produced in the world was published in the Soviet Union, and the Soviet publishing industry was translating more titles than any other country (Childs & McNeil, 1986, pp. 200–204). Ethiopia provides an interesting example of how the Soviets used book distribution to increase its influence in emerging countries. In l973, the year before Emperor Haile Selassie was overthrown and replaced by a Marxist government, the Soviet Union did not publish a single title in Amharic. By 1976 it had published fifteen titles in Amharic in 300,000 copies, and by 1980 twenty-four titles in Amharic in 820,000 copies, which probably represented over a third of the world's Amharic book production for that year (Freeman & Righetti, 1984, p. 31). The overseas distribution of publications included extremely low-cost or free issues of the more than seventy abstracting journals published by VINITI in Moscow, with their reviews of scientific and technical articles in sixty-five languages.

Geopolitical Impact of the Soviets' Information Empire

To understand fully the implications of the rise and fall of Soviet influence on the information professions, we have also to take into consideration American reactions to the widening Soviet influence in this sphere from the 1950s on. Evidence of American concern about the presumed efficiency of Soviet technical information abstracting and dissemination dates from even before the *Sputnik* launching and continues right through the 1980s. In 1956 Jesse Shera, dean of the library school at what became Case Western Reserve University, focused on the threat of Soviet information hegemony in his keynote speech to the Special Library Association. He bemoaned the lack of progress in American information technology and called attention to Soviet advances in large-scale abstracting. "What new bibliographic achievements have we to show since the UNESCO Conference on Improving Bibliographic Services met in Paris in 1950?" Shera asked his audience. He warned that "there is more to concern us here than a mere decline in national prestige. On our own ability to put knowledge to work may rest the very future of our civilization and the perpetuation of our cherished way of life. We are engaged in a grim game; we may not long hold all the high cards, if indeed we do now and—make no mistake about it—this time we are playing for keeps" (Shera, 1965, p. 61). When Eugene Garfield launched his new Institute for Scientific Information in 1960, he called it a "free enterprise alternative" to VINITI (Garfield, 1960, p. 198), implying that ISI would do for American science what VINITI presumably had done for Soviet science, that is, enhance productivity. Anxiety about VINITI continued through the 1980s: In 1981 MIT convened a conference on Soviet abstracting addressed by George Vladutz, a former VINITI official who had emigrated to the United States and who worked for ISI (Vladutz, 1981).

A full discussion of the American overseas response to the Soviet rivalry in information professionalism during the Cold War lies outside the scope of this paper. The most cursory review of the American library literature of the period, however, shows that American anxieties such as those expressed by Shera were translated into tangible aid programs to stimulate the flow of American technical know-how to parts of the world vulnerable to communist influence. Just a few such programs were the book program for Indonesia starting in 1964, sponsored by the United States National Academy of Sciences; the distribution of twenty million textbooks to the students of the Philippines in the early 1960s, funded by the United States Agency for International Development (USAID); and the use of USAID and Public Law 480 funds to construct the University of Mindanao in the Philippines in 1968 (Kaser, Stone, & Byrd, 1969). In addition, between 1950 and 1962, the United States Information Agency (USIA) financed the publication or translation in English of 13,632 titles in 123,969,405 copies for distribution overseas (Elder, 1968, p. 265). During the post-Vietnam era, the United States reduced the scale of its international initiatives (e.g., withdrawing from UNESCO in 1984). Competition between the Soviet Union and the United States for world ideological leadership nonetheless acted as a brake on American withdrawal from some overseas assistance programs, especially after the first Reagan administration made ascendancy over the "evil empire" an administration priority.

Conclusion

All of the Soviet Union's international assistance programs in information infrastructure development have now stopped. In 1998 the Krupskaia Institute of Culture (now the St. Petersburg Academy of Culture) will graduate its last Moscow-subsidized foreign students. Together with the possibilities of free higher and continuing education, the subsidized flow of scientific and technical information from Moscow to its former client countries has also stopped, as Russia's publishers struggle to survive in a market economy. At this date, the only former members of the Soviet bloc that have substantial access to the world's current scientific and technical information are those such as Poland, Hungary, and the Czech Republic, which have the hard currency to pay for it.

At the same time the non-Soviet agencies that subsidized information to hard-currency–poor countries in the 1970s and 1980s have radically diminished their assistance. UNESCO has been downsizing since the American and British withdrawal in 1984, and the once-lavish book distribution programs of USAID and the USIA have shrunk dramatically. The United States, no longer competing with the U.S.S.R. for the affections of the nonaligned developing world, has shifted its focus to influencing Russia itself. Since 1994, under the aegis of the Freedom Support Act, the USIA has been subsidizing the education of scores of students from the former Soviet Union in American library and information science. Support under this act is not available to students from the former Soviet client states.

Meanwhile, the developing world is littered with centralized, government-operated information centers operating in a virtual vacuum ever since the disappearance of the Soviet information supply upon which they depended. This vacuum can be expected to continue far into the twenty-first century unless another substitute for the market system is found to replace the Soviet Union's assistance programs. For over thirty years they offered a window on the international world of science and technology to countries unable to pay the market price for such access. The officials of a number of developing countries remain convinced today that the Soviet system offered greater advantages to emerging countries than does international capitalism (Bui, 1997, p. 102).

Acknowledgments

The author wishes to express her gratitude to the U.S. Fulbright Commission for making it possible to do the research for this article while she was a 1997 Fulbright Scholar to Russia.

References

Barghoorn, F. (1960). *The Soviet cultural offensive: The role of cultural diplomacy in Soviet foreign policy.* Princeton, NJ: Princeton University Press.

Bergen, D. P. (1962). Communist and American cultural strategy in Asia, Africa, and Latin America. *Library Quarterly,* 111–132.

Bui, L. T. (1997). Sovremennoe sostoianie oblastnikh bibliotek vo V'etname i rossiskii opyt organizatsii i bedeniia nauchno-issledovatel'skoi raboty [The current status of regional libraries in Vietnam and Russian experience with the organizing and management of scholarly and scientific research]. *Problemy natsionalnykh bibliotek i regional'nikh bibliotechnyikh tsentrov.* St. Petersburg: Natsional'naia Biblioteka Rossii.

Childs, W. M., & McNeil, D. E. (1986). *American books abroad: Towards a national policy*. Washington, DC: Helen Dwight Reid Educational Foundation.

Dissertations (1997). Catalog of dissertations in the Department of Library Science. Archives of the St. Petersburg Academy of Culture.

Elder, R. E. (1968). *The information machine: The United States Information Agency and American foreign policy.* Syracuse: Syracuse University Press.

Freeman, W. E., & Righetti, S.(1984). *Soviet book exports 1973–82*. Washington, DC: USIA Department of Research.

Garfield, E. (1960, November). Institute for Scientific Information created. *American Documentation,* 198.

Gorbacheva, I. G. (1981). *Sotrudnichestvo evropeiskikh sotsialisticheskikh stran v realizatsii programmy universal'nogo bibliograficheskogo ucheta*. [Collaboration of the European socialist countries in the realization of universal bibliographic control]. Moskva: Moskovskii gosudarstvennyi institut kul'tury.

Graham, L. W. (1975). The development of science policy in the Soviet Union. In T. D. Long & C. Wright (Eds.), *Science policies of industrialized nations* (pp. 12–15). New York: Praeger Press.

Kaser, D., Stone, C., & Byrd, C. K. (1969). *Library development in eight Asian countries*. Metuchen, NJ: Scarecrow.

Kozlovskii, B. (1948a, May). Burzhuaznye biblioteki na sluzhbe reaktsii. *Bibliotekar',* 27–35.

Kozlovskii, B.(1948b, June). Burzhuaznye biblioteki na sluzhbe reaktsii. *Bibliotekar',* 39–42.

Kneen, P. (1985). *Soviet scientists and the state.* Albany, NY: State University of New York Press.

Le Riverend, Kh. (1982). Leninskie printsipy bibliotechnogo dela i biblioteki sotsialisticheskoi Kuby [Lenin's principles of librarianship and libraries in socialist Cuba]. *Bibliotekovedenie i bibliografiia za rube-zhom,* 3–7.

Mezhdunarodnyi Tsentr nauchnoi i tekhnicheskoi informatsii (1977). *Podgotovka i Pobyshenie Kvalifikatsii informatsionnykh rabotnkikov v SSSR.* Moskva.

Mikhailov, A. I., Chernyi, A. I.,& Giliarevskii, R. (1984). *Scientific communication and informatics.* Robert H. Burger, (Trans.). Arlington, VA: Information Resources Press.

Pudov, Kh. (1982). *Soveshchanie tsentrov nauchno-issledovatel'skoii metodicheskoi raboty sotsialisticheskikh stran v oblasti bibliotechnogo dela: materialy* [Meeting of the research centers for library methodology of the socialist countries]. Moskva: G. B. i. Lenina.

Rasizm v deistvii. [Racism in Action] (1955). *Bibliotekar',* February, 60–61.

Raymond, B. (1979). *Krupskaia and Soviet Russian librarianship 1917–1939*. Metuchen, NJ: Scarecrow.

Richards, P. S. (1992). Education for information science in the Soviet Union. In M. Williams (Ed.), *Annual review of information science and technology* (pp.267–290). Medford, NJ: Learned Information.

Shera, J. (1965). *Libraries and the organization of knowledge.* Hamden, CT: Archon Books.

Sherry, M. (1995). *In the shadow of war: The United States since the 1930s*. New Haven, CT: Yale University Press.

Tvardovskaia, N. I. (1984). Nauchno-prakticheskaia konferentsiia 'bibliotechnoe delo kapitalisticheskikh stran i ideologicheskaia borba na sovremennom etape' [Scholarly/practical conference on the current status of librarianship in capitalist countries and of the ideological struggle. *Bibliotekovedeniei bibliografiia za rubezhom,* 68–79.

U.S. Department of State (1987). *Soviet influence activities: A report on active measures and propaganda 1986–1987.* Washington, DC: Department of State Publication 9627.

Varaksina, R. IA. (1976). Seminar bibliotechnikh rabotnikov stran Azii, Afriki i Latinskoi Ameriki 'Planirovanie natsional'nikh sistem'. Moskva/Alma Ata. [Seminar of librarians from Asia, Africa and Latin America on planning national systems, held in Moscow and Alma Ata]. *Bibliotekovedenie i Bibliografiia za rubezom,* 72–94.

Vladutz, G. (1981). The All-Union Institute for Scientific and Technical Information (VINITI). Cambridge, MA: Program on Science, Technology and Society, Massachusetts Institute of Technology; Russian Research Center, Harvard University.

Vucinich, A. (1984). *Empire of knowledge: The Academy of Sciences of the USSR 1917–1970.* Berkeley, CA: University of California Press.

Restoration of Japanese Academic Libraries and Development of Library and Information Science: The Contributions of Shigenori Baba

Takashi Satoh

Abstract

World War II essentially destroyed the Japanese social system. The academic world also suffered great losses. One of the officials who worked for academic reconstruction was Shigenori Baba (1909–1993). Born of a wealthy family in Tokyo, Baba enjoyed reading classics and learning foreign languages. He was interested in Christianity and was baptized in his youth. Baba majored in electrical engineering in college with studies in physics, and he worked as a researcher for the Furukawa financial combine, where he produced a large number of scientific abstracts. As his achievement was recognized by the government, he began to work, in the last stage of the war, for a governmental organization that controlled all scientific information. After the war Baba worked for the Ministry of Education, where his achievements were extensive: developing a union catalog, founding district library conferences, improving university librarians' status, making dictionaries of technical terms, and introducing the idea of documentation to university libraries. An academic as well as a government executive official, Baba was sent to Europe to study documentation, which led to his theory of documentation based extensively on mathematics. As a professor for the National College of Library Science, Baba's major accomplishments were bibliographical analysis by quantitative methods that became the present bibliometrics; his unique bibliographical equation; the "Union Index"; transliteration of non-Roman scripts; and dictionaries of technical terms in eight languages, including Czech entries.

Introduction

Shigenori Baba, who passed away in 1993, contributed greatly to the reconstruction of Japan's academic libraries and the revival of its devastated postwar world of learning. Baba's special achievement, among others, is the development of library and information science through the theory and practices of documentation. As the author of 21 books and more than 130 papers, Baba's collection of books, records of meetings of the committees he took part in, letters, and research notes are plentiful. Now, as most of the material has been classified, I wish to show some of Baba's achievements over fifty years.

Educational Background

Shigenori Baba was born in 1909 to a wealthy family of a banker, descended from a dignitary of the Tokugawa shogunate in the Edo era. His father was a learned man and wanted his children to be highly educated. They were taught English and French privately by Cambridge-educated and other scholarly teachers. It was common for boys from former samurai families to learn Chinese classics, so Baba's father invited Dr. Shionoya, a reputed scholar of Chinese classics, to teach his son. As a boy Baba took an interest in many academic fields, including classic literature from both the East and the West and natural science. He was also interested in Christianity and was baptized at age fifteen at the Fujimicho Presbyterian Church in Tokyo by the Reverend Masahisa Uemura, the founder of the church and a well-known theologian.

Baba chose archaeology as his major and entered Waseda University, expecting to study under Shigeyasu Tokunaga, who was then a prominent archaeologist. However, the great stock market crash in 1929 interrupted his studies. Because his father's bank was in financial difficulties, Baba was forced to choose a more practical field of study. He transferred to the faculty of electrical engineering where a fellow Christian at the church, Tadaoki Yamamoto, was teaching. Later Baba married Yamamoto's second daughter. As he was fond of mathematics and physics, he found no difficulty in studying electrical engineering. In his graduation thesis he focused on what is known as frequency increase and wrote it in German under the guidance of F. Niethammer of Vienna University.

Showing a talent for learning foreign languages, which was partly because of his constant visits to the Christian church, Baba became well versed in European languages, including Greek and Latin. It was also easy for him to write papers in German. He was interested in languages around the world and later engaged in the transliteration of Asian languages. It was quite natural that he preferred basic theoretical study rather than practical electrical engineering. Because Baba did not know much about how electricity was used in actual sites, his professors suggested on-the-job training for several years and advised him to join the Furukawa Electric Comapny, a member of the Furukawa financial combine.

The First Steps

Baba joined Furukawa in 1934, where he was assigned much practical work even though he was posted as a researcher. There he recognized how important it was to have theoretical knowledge to cope with actual problems and to verify the theory by reading electrical literature from many parts of the world. These efforts increased his electrical engineering knowledge so much that he was more informed than most scientists. Even professors asked him for advice based on his knowledge.

There was a library at Furukawa, but there was no provision of information service. Baba wanted to set up a question-answering system there. He made use of sections A and B of *Science Abstracts*, which he had been accustomed to doing while a student, as well as primary journals. He initiated an in-house abstracting service: a card catalog of bibliographic data with abstracts of relevant literatures that grew day by day. He also analyzed the types of columns that made up each journal, such as *Archiv für Elektrotechnik*, *Electrotecnica*, *Elektrotechnik und Maschinenbau*, *Elektrotechnische Zeitschrift*, *Journal of the Institution of Electrical Engineers*, *Revue Générale de l'Électricité*, and so on.

His card catalog was well organized; items were arranged according to subject fields and types of literature, making it easy for him to compile his "Bibliography of Literatures on Electric Wires and Cables" and distribute it to all the research sections.

During his ten years at Furukawa he had made 75,000 abstracts from books, papers, patents, and technical reports (Baba, 1977, pp. 4–5). This number is even larger than the 3,380 abstracts and 920 book reviews completed by Wilhelm Ostwald over a seventeen-year period (Satoh, 1987). Later in his life Baba told me that he had read Ostwald's books carefully. These two had a great deal in common. When Ostwald heard that scientists often found such work as writing textbooks, abstracts, and book reviews troublesome, he wrote that textbooks should be written by the first-class scientists and that scientists would neglect their duty if they avoided what seemed to give them trouble. One of the great pleasures of Shigenori Baba was to give his knowledge to those who needed it (Baba, 1977, pp. 6–7). At Furukawa he worked as a researcher, but he cultivated his ability as a documentalist, even though the word documentation was not well known at the time.

War Assignment

Baba's work of making high-quality bibliographies and guides to the scientific literatures attracted considerable attention from government officials. When Japan went to war, the government needed to control all science and technology to keep productivity high. Baba was asked to work as a science officer for the Board of Technology in 1943, when the war situation became even worse. His first task was to categorize research papers for scientific mobilization. Here again, abstracting services were essential to the final judgment for selection, and Universal Decimal Classification (UDC) class numbers were assigned to each item on the eighty-column punched card. The UDC was privately translated from the French edition published in Belgium in 1939. The work was essentially a private directory of scientists in Japan for the use of government officials.

When the Japanese surrendered in 1945, the whole nation fell into ruin.

Executive Official

In 1945 the thirty-six-year-old Baba began work with the Ministry of Education. From this post he began his twenty-year effort to restore the Japanese academic library system.

Baba believed that literature was the foundation of research and that research was the basis of national power. Hantaro Nagaoka, the world-famous physicist and one of Baba's acquaintances, said, "The post-war restoration begins with study and learning. I want to have the literature in Japan reorganized and a union catalog made so that everyone may use it" (Baba, 1977, p. 10). The work was much more difficult than expected.

Sociologist Chie Nakane (1978) observes that Japanese society is vertically divided as is the academic society of universities. There are not enough contacts among large and small libraries even within one university, still less with libraries of other universities. A library conference was founded in the late 1920s as a liaison organization, but the conference was divided vertically into two bodies: one for the imperial universities and one for private universities. These circumstances made it difficult to produce a union catalog to connect all the libraries, although in 1931 the Science Council had tried building a union catalog of scientific magazines.

Just after the war ended, the shortages of personnel, writing paper, pens, and funds made the work even more difficult. The biggest problem, however, was to find the right balance between the policy of the Japanese government and that of the occupation forces. The occupation forces aimed to dissolve the Japanese military and to build a democratic social system. For information sources their plan was to replace the old Imperial Library with a National Diet Library like the U.S. Library of Congress and to transfer the task of making the union catalog from the Ministry of Education to the newly built Diet Library. Baba thought that the union catalog was indispensable to the policymaking of the ministry and insisted that the union catalog should belong not to the Diet Library, which is a legislative organ, but to the ministry, which is an administrative one (Baba, 1977, p. 10).

Professor Akira Nemoto (personal communication, 23 May 1997) of Tokyo University, who recently investigated GHQ documents told me that Baba's name was often mentioned in them. Baba had a heated dispute with Robert Downs and other officials of the Occupation Forces GHQ over this issue. The union catalog, while he was away from Japan, was placed under the control of the Diet Library. But now it is maintained by the National Center for Science Information Systems of the Ministry of Education.

Baba was able to organize all the Japanese universities into ten district groups. Each district had its main library equipped with a microcopy machine, which was valuable at the time when copying apparatus was not in common use. Each district started a journal that made it possible to exchange opinions among the members. The journal carried contributed papers regularly, which gradually improved the status of the library personnel in society. What was important was that every university, regardless of national, private, prefectural, or municipal status, could join the district library conferences. Since the tendency in Japanese culture is to regard something made by a national authority as higher than something made by a private institution, this development was unusual.

Library workers were not highly respected in the academic world, and no rules or standards to employ them had been established. To work for a library of a national university, applicants needed to take the civil service examination. Baba negotiated with the National Personnel Authority about the matter and gained a new category of library science in the examination to select the top-level executive officials.

For academic restoration, Baba organized the Branch Council for Technical Terms in the Ministry of Education and invited leading scholars of each field to standardize the technical terms. Like Ostwald, he believed that the world of study and learning must be furnished with standardization. (He later lamented that this opportunity of standardization once brought about serious antagonism among several Buddhist denominations while they were discussing one word in the field of religion.) Thus far, twenty-seven dictionaries of technical terms have been published, including a dictionary of library science, in which Baba played a big role (Ministry of Education, 1958); the revised edition appeared in 1997.

Study in Europe and Education in Japan

Baba's twenty-year tenure at the Ministry of Education was greatly enriched by his one year of documentation study in Europe. In 1952 he received a six-month UNESCO fellowship, and the ministry added an additional six months of study time. In Japan the term *documentation* was unknown, and yet this was what Baba had been doing. He visited the Netherlands, Belgium, France, Switzerland, West Germany, Denmark, Finland, Sweden, England, and Spain. Among the fellow researchers and friends with whom he enjoyed long-time friendship were Frits Donker Duyvis of the International Federation for Information and Documentation, Julien Cain of the Bibliothèque Nationale, Paule Salvan, Otto Frank, Hanns Eppelsheimer, Frank Francis of the British Museum, and Jose M. Albareda Herrera. He also had good relationships with Ralph Shaw and Foster Mohrhardt of the Library of the U.S. Department of Agriculture, George Bonn of the University of Illinois, and Vivian Edmiston. These Americans, who were sent to Japan by U.S. government agencies, recognized Baba's ability to resolve difficult matters.

In Europe, Baba visited historic libraries and documentation facilities and worked out a detailed plan for his theory of documentation. He understood that universal philanthropy forms the basis of documentation, which shows each nation's strength. In this sense he insisted that documentation be done by natives. The word *documentation* was soon to be replaced by such words as *information science* or *informatics*. Baba, however, continued to use the term, apparently feeling that it carried a sense of European culture.

After his study in Europe the level of Japanese libraries and information science seemed quite low to Baba. He thought that Japanese libraries and information science were primarily oriented toward public libraries rather than research libraries, which were badly needed to meet researchers' demands. This led him to plan documentation workshops in all the districts in which he had organized library conferences. These workshops stressed the concepts, theories, and technical skills of documentation.

After twenty-three years' service to the government he was invited to become a professor at the National College of Library Science in 1965, where he lectured and trained many librarians who now form the backbone of local libraries throughout the country. In 1971 the college began a Documentation and Information Science course; it was the outcome of his long-term efforts, and Baba was appointed the head of the new course.

Contributions to Library and Information Science

Baba called the field of his study *literoscience* (Baba, 1971a), but he also insisted on the use of the term *documentation* in the course. His students published a 272-page book of his collected papers to commemorate his retirement (Baba, 1977). The first half of the book contained an interview with him, and eleven papers of his own selection made up the rest. He wanted these eleven papers classified into the following three parts.

Part 1: Basic Concept of Documentation

1) Essays on documentation, in Japanese, September and October 1962.

Part 2: From Collection of Information to Collection Building

2) Quantitative method of selecting literature, in Japanese, May 1958.
3) Several facts pertaining to Literostructure: from the viewpoint of comparative scientific material, comparative library science (including comparative studies in cataloging and catalogs, etc.), and library linguistics, in Japanese, March 1967.
4) An aspect regarding quantification method for selection of bibliographic vessels, in English, 1965.
5) Fundamental theory of bibliographical structure for collection development, in Japanese, February 1971.
6) Unchanged and transformed contents in the higher-ordered literature columns of learned periodicals, in Japanese, June 1971.
7) Literature science (Literoscience)—Literometrics as its quantitative base—an example from the subject "cataloging and catalog," in Japanese, March 1975.

Part 3: Bibliostructoanalysis and Union Indexing

8) Fundamental theory of comparative bibliographical science, with special reference to bibliographical materials, in Japanese, March 1973.
9) Bibliographical science for documentation activities: fundamental theory of comparative science, especially regarding bibliographical characteristics of bibliographical materials, in Japanese, 1972.
10) Indexing of (subject) indexes: union indexing and index, in Japanese, March 1969.
11) Union index of books in the field of library automation published in the United Kingdom and the United States, in Japanese, October 1974.

Baba put literoscience and its outcomes into shape and tried to make a theory of it. His papers were written in a crabbed Japanese style that was greatly influenced by Chinese classics. Thus his papers were often criticized as too difficult to understand, too abstract, and sometimes impractical.

Nonetheless, his method of study has three characteristics. First, he had a quantitative method, as shown in his papers in part 2. He already had reached the idea of core journals (a kind of *Kernliteratur* in his German), using early citation analysis techniques. These methods are known as bibliometrics today. Reading his papers closely reveals uncommon concepts, including segregation and scattering. Second, his idea of "bibliographical equation" is unique. He analyzes the structure of literature thoroughly as a research object. He notes literature as $N = D + L + B + I + \Xi$.

He calls D a "bibliographical identifier" in which d_1 is, for instance, the author's name, d_2 is the title of a book, . . . so $D = D_1 + D_2 + D_3 + \ldots + D_i$. L is a "semantifier" and l is made from many semantic factors, so $L = l_1 + l_2 + l_3 + \ldots + l_i$. B is a "bibliographical referencifier." I is a "bibliographical indexifier," and Ξ equals "collatifier" in which tables, diagrams, and graphs in a literature are included. Finally, Baba (1971a) notes a general equation of literature as:

$$N^{o_i e_i k_i \alpha_i} = D^{o_i e_i k_i \alpha_i} + L^{o_i e_i k_i \alpha_i} + B^{o_i e_i k_i \alpha_i} + I^{o_i e_i k_i \alpha_i} + \Xi^{o_i e_i k_i \alpha_i}$$

(o = order; e = element; k = kind; α = viewpoint)

With this basic equation he considered fundamental theories of selection of literature, bibliographical participation, and so on. The bibliographical equation is the ultimate goal of his methodology.

Third, he maintained a sophisticated card catalog. Baba took notes on books he read on 3 × 5 cards. On each card he wrote what he knew and its source and linked related items based on his analysis. He did not hesitate to write down his sources, though he had a good memory. Among the enormous collection of cards he left, there are substitute cards cut during the postwar years from used paper. His card catalog could be compared with Ostwald's "Kartothek." We may call it hypertext today.

The method of analyzing and synthesizing knowledge grew to become the "Union Index." He published a bulky work of 657 pages as an index to books in the field of documentation written in English (Baba, 1970). The book is the fruit of a labor of love. He analyzed the content of some sixty books in the field, made detailed indexes to them, added citation links and related terms among them, and finally unified them in one volume. This index represents an amazing amount of work, since it was all done manually using cards.

Transliteration and Polyglot Dictionaries

Finally, his contribution to the study of languages must be acknowledged. Baba published a 433-page book on the transliteration of non-Roman characters into Roman characters (Baba, 1968). The book gave transliterated alphabetic lists of some 150 non-Roman languages, based on International Standardization Organization principles, and was supplemented for the first time with Japanese equivalents to Bulgarian, Czech, and Greek characters. He continued to publish papers on the transliteration of Indian, Indonesian, Finno-Ugrian, Baltic, and other languages (Baba, 1971b, 1973, 1975a, 1975b).

An interesting work that combines Baba's study of languages with bibliostructure analysis appeared in 1961 (Baba, 1961). This bibliography was a useful guide to those who could not read Russian publications, which became of great significance in the fields of science and technology. It was an annotated bibliography of 122 pages, but the structured annotation under each source gave such guidance as the method of service (i.e., translation, abstract, index, or others), the type of column that offered the service, the languages used, the publishing country, coverage of Russian literature, and so on. The bibliography listed various sources from twenty countries.

After his retirement in 1977, Baba kept compiling a set of polyglot dictionaries in the field of library and information science. Tens of thousands of cards were being prepared, which covered in total English, German, Dutch, French, Spanish, Italian, Danish, Finnish, Swedish, Norwegian, Russian, Czech, Polish, Rumanian, Bulgarian, Hungarian, Serbo-Croatian, Albanian, Greek, and Japanese technical terms. He completed a notebook of terminology with Czech entries, each of which had English, French, German, Russian, Polish, Spanish, and Japanese equivalents. While he was negotiating for its publication, Baba died of a heart attack at the age of 84 early in 1993. A draft of the terminology was left open on his desk.

Conclusion

He was a man of vigor throughout his life. In his later years, however, his arms suffered from tendonitis caused by lifting card catalog trays, which contained hundreds of thousands of cards. Shigenori Baba, who worked as a documentalist, science officer, executive official, educator, lexicographer, linguist, and above all, philanthropist, accomplished valuable works. From his lectures and writings we can see his vision of documentation as standardization in technical aspects on the one hand and universal philanthropy in spiritual aspect on the other.

Acknowledgments

I am grateful to Keiichi Kawamura, librarian of Dokkyo University School of Medicine, for his kind and thoughtful assistance in preparing the manuscript.

References

Baba, S. (1961). *Roshiago bunken no kansetsu riyoho* [Method of indirect use of Russian literatures]. Tokyo: Gakujutsu Bunken Fukyukai.

Baba, S. (1968). *Hi-romaji no romaji (shoshiteki) honji* [Transliteration of non-Roman characters into Roman characters]. Tokyo: Gakujutsu Bunken Fukyukai.

Baba, S. (1970). *Union index of books in the field of documentation*. Tokyo: Gakujutsu Bunken Fukyukai.

Baba, S. (1971a). Zosho kosei ni kansuru ichi-ni no chiken. Dai sanpo: Zosho kosei sanka kiso riron oyobi zoku sogo-sakuinka to sogo-sakuin [Several known facts pertaining to literostructure. Third report: Fundamental theory of bibliographical structuring participation and union indexing and union index—continued]. *Memoirs of the Toshokan Tanki Daigaku, 4*, 63–129.

Baba, S. (1971b). Hi-romaji gengomoji to sono honji: Kochishina; Kipurosu onsetsumoji sono hoka no teisei hosoku [Transliteration of non-Roman scripts into Roman scripts: Scripts of Cochin-China and supplements and corrections to scripts of Cyprus and others]. *Memoirs of the Toshokan Tanki Daigaku, 5*, 67–94.

Baba, S. (1973). Hi-romaji gengomoji to sono honji: Indo [Transliteration of scripts in Indian into Roman scripts]. *Memoirs of the Toshokan Tanki Daigaku, 7*, 3–29.

Baba, S. (1975a). Hi-romaji gengomoji to sono honji: Indoneshia [Transliteration of some Indonesian non-Roman scripts into Roman scripts]. *Memoirs of the Toshokan Tanki Daigaku, 9*, 77–97.

Baba, S. (1975b). Romansu, Geruman, Surabu, Fino-Ugoru, Baruto, Ruhn shomoji to honji aruiwa tensha [Romance, Germanic, Slavonic, Finno-Ugrian, Baltic and Runic scripts and their transliteration or transcriptions]. *Memoirs of the Toshokan Tanki Daigaku, 10*, 47–58.

Baba, S. (1977). *Dokyumenteshon—Baba Shigenori sensei ronbunshu* [Documentation—collected papers of Shigenori Baba]. Tokyo: Publishing Committee of Collected Papers of Prof. Shigenori Baba, National College of Library Science.

Ministry of Education. (1958). *Monbusho gakujutsu yogoshu: Toshokangaku-hen* [Japanese scientific terms: Library science]. Tokyo: Dainippon Tosho.

Nakane, C. (1978). *Tateshakai no rikigaku* [Mechanics of a vertical society]. Tokyo: Kodansha.

Satoh, T. (1987). The bridge movement in Munich and Ostwald's treatise on the organization of knowledge. *Libri, 37*, 1–24.

History Reviewed and Revisited

History Review: The Development of Information Science in the United States

Robert M. Hayes

Introduction

This paper expands on the summary provided at the conclusion of the conference on the history of science information and information science. It provides a brief review of the history of developments in those fields in the United States, including consideration of some contextual dimensions within which those developments should be seen. I am going to discuss seven timelines spanning the past fifty years, divided into three time periods: 1948–64, 1964–80, and 1980–98. I conclude with predictions about the coming five-year period. The seven timelines are the context of scientific, social, economic, and political developments; the context of commercial, industrial, and consumer information resources; development of the underlying technologies; the context of publishing; the context of library concerns; development of library automation; and development of information science and services. These timelines all interact with each other, and I will first try to highlight some of their interactions. Of special importance were developments in the commercial, industrial, and consumer sectors of parallels to those in the library and information science sectors. In many respects, though almost unheralded, the development of automation in libraries demonstrated the feasibility of its application in the larger arena. I think especially of the leadership of the National Library of Medicine in proving the feasibility of high-quality typographic output from computers, thus laying the groundwork for computer-controlled photocomposition in publishing. I think of the development of online reference data services, first demonstrated in libraries and then finding great acceptance in industry. I think of the distribution of CD-ROMs, which found their first real market in libraries. As I proceed through this history, these interactions should be evident.

Also of special importance has been the effect of technological capabilities not only on what could be done but also on the very perception of how to do it. This was evident in the early years, when the limits of punched-card technology as well as the means for the logical processing it embodied largely determined how people thought about using technology. Since then the capabilities of the technology have grown so dramatically that an understanding of how to use it in information science continually lags behind.

But perhaps the most important of the interactions were the effects of the scientific, social, economic, and political context. They provided and continue to provide the rationale for commitment of resources to all of the developments presented here.

I will interlace into the discussion some elements related to my own professional involvement because it represents the perspective I bring to the developments and to some extent my own participation in them. In this way you can judge what I say in the light of my experiences.

1948 to 1964

I came to this field during the period from 1948 to 1964. At the beginning of it, after receiving my Ph.D. in mathematics from UCLA, I started in systems work and the use of computers. In the middle of the period I gained some knowledge of applications to information retrieval. During the last third I formed my own company and by the end of the period had joined the UCLA School of Library Service.

From the period when Vannevar Bush was the science adviser to the president, the needs for information in science and technology provided a continuing rationale for development of automated information management systems. Indeed, this period started with Bush's frequently cited article, "As We May Think" (Bush, 1945a). In a very real sense it foretold virtually everything we have since seen in terms of development in our field. Much though that article has been cited, however, a far more important document is the earlier 1945 report of the Office of Scientific Research and Development (OSRD) of which Bush was chair. It urged continued support for scientific research after World War II ended. By doing so, it led to one of the most important steps in the history of science in the United States: It proposed what in 1950 became the National Science Foundation and the National Institutes of Health (Bush, 1945b).

Of specific relevance to science information and information science, the OSRD report identified many of the continuing themes that have been important to our field: international exchange of scientific information, publication of scientific information, lifting of security restrictions for a broad dissemination of scientific information, encouragement of publication, and library aids. In the discussion of library aids it states, "Adequate technical libraries are an indispensable tool for research workers." Then it makes specific reference to information technology: "It seems probable that use of cataloging and sorting devices now available in the form of business machines and microfilm technique might go far to improve present methods of searching the literature and making bibliographies" (Bush, 1945b, pp. 112–115). Indeed, by the end of the period from 1948 to 1964, the realization of that conjecture was well under way. Let us examine each of the seven timelines during that period.

The Scientific, Social, Economic, and Political Contexts

During this period the United States was the dominant economic power in the world and the major producer of scientific research publications. The reason is evident: It was the only major country that had not suffered catastrophic losses during World War II. Its economic infrastructure was intact, as was its academic superstructure. Fortunately, I think, it used its economic power wisely, as best represented by the Marshall Plan that made U.S. resources available for the rebuilding of Europe and the Far East. From the standpoint of science information, at the beginning of this period (1948–50), the United States produced more than 50 percent of the world's scientific publications.

This was the period of the Cold War, which reached its peak during the Korean War and the Cuban missile crisis. One of the most critical events came in the middle of this period—the launching of *Sputnik* by the Soviets in 1957. This event shocked the military, industrial, and scientific establishment of the United States and led directly to some of the most crucial developments in our own field.

Even before those epoch-making events, though, there was a continuing emphasis on the need for information services to support U.S. national defense. In the early days the needs lay in the management of the flood of documents taken from Germany after World War II; this "documentation" clearly required development of new tools to support that management and the related tasks in storing and retrieving data from such massive amounts of material. Later, though, as the Cold War became more intense, the intelligence community—the Central Intelligence Agency (CIA), National Security Agency (NSA), and the intelligence arms of each of the branches of the Armed Services—required means for assembling, storing and retrieving, and analyzing even more massive amounts of data. Their needs more than any others supported the development of computer hardware and software and led to the development of "systems work" as the means for assessing requirements and implementing efficient systems (*Documentation, indexing, and retrieval*, 1961).[1] The operational arms of branches of the Department of Defense, especially the Air Force, also required the development and implementation of systems that would support their operating requirements. The "early-warning" systems of the North American Aerospace Command (NORAD) and the Strategic Air Command (SAC) and the tactical systems to support troops in the field both led to significant investments in information technologies by each of the branches. They also led to commitments of the major aerospace companies to work in the field of information technology. Indeed, the Air Force and the Office of Na-

[1] In *Documentation, indexing, and retrieval*, see especially pp. 63–64 (CIA Intellofax system) and p. 65 (Minicard system).

val Research were among the most important sources of funding for research in information science.

That was the context for the recommendations concerning science information in the Report to the President and for Bush's "As We May Think." But the demands reached a crescendo as a result of *Sputnik*. The science adviser to the president became even more an advocate for science information. The National Science Foundation launched its Office of Science Information Service to support development and implementation of automated systems and to sponsor the necessary research; of special importance was its funding of the automation of Chemical Abstracts Services. It also published several summaries of both research activities and operational services ("Current research and development," 1957–1969; "Nonconventional scientific and technical information," 1958–1966; "Specialized science information services," 1961).

I have always regarded one development in the U.S. military during this period as an important application of information science, but one which the field has largely ignored. It was the Federal Parts Cataloging program, which started at the very beginning of this period and became the basis for inventory management throughout the federal government and, indeed, in NATO activities as well as in the U.S. military (Hayes, 1992).

Further, consistent with the recommendations of the OSRD, health care and related research became a priority of the federal government. The National Institutes of Health were established and the Surgeon General's Library became the National Library of Medicine. Toward the end of this period the National Library of Medicine launched its efforts to automate the production of *Index Medicus* (Miles, 1982).[2]

Commercial, Industrial, and Consumer Contexts

In this period there was limited recognition of the role of information in our society. Its use within companies was primarily for the purpose of internal accounting, and the amount of data involved was small. The number of transactions handled, while large in terms of the processing capabilities of the time, was tiny compared with the situation today. There was at best limited use, if any, of external information, although industrial special libraries did serve as a means for access to it.

I had a discussion in 1960 with the processing staff of Bache & Company (then one of the major stock brokerage firms on Wall Street). I was a consultant to them because they were implementing one of the first computer systems in such a context. They proudly announced that they had designed the system to handle three times the workload expected during a then-typical day on Wall Street, in which the volume of activity was three million shares a day. I asked what they would do when the volume of activity increased to ten million shares per day, and they looked at me as though I were crazy. That level, however, was reached within seven years and today is consistently over five hundred million shares per day.

For the consumer, information was a little-recognized commodity. At the beginning of the period movies were the dominant means of entertainment, though television was just beginning to pervade our lives. (I vividly recall watching, through store windows, the telecasts of the McCarthy hearings and especially the devastating assessments of Edward R. Murrow and of Judge Welch.) Further, sales of books, newspapers, and magazines in the United States were at about $1.4 billion a year—$7.67 per capita.

Information Technology

It is with almost shock that I recall the nature of the technology during that seventeen-year period and compare it with what we have today. In the beginning data processing meant punched cards and key-operated accounting machines, with punched tape (like Teletype tape) as the "common language" for communication among machines. Computers were limited in capabilities, in numbers, and in applications. I worked at UCLA on one of the first of them, the Bureau of Standards Western Automatic Calculator (SWAC). While large and fast for the time, its capabilities were exceptionally limited. Today I can hold in my hand greater computing power, greater functionality, and greater ability to apply it and use it—and it all runs on two AA batteries!

The first small-scale computers for application in business and similar operations (among which I would include libraries and information services) appeared in the early 1950s, and larger systems—mainframe computers—became widespread by the end of the period. But all of that equipment suffered from the lack of adequate means for input, storage, display, and output of data. Punched cards and punched tape were the only means for input, and they operated at data rates roughly equivalent to ten characters per second—the speed of

[2] In Miles, see especially p. 365 (support grant from Council on Library Resources in 1958), pp. 372–373 (GRACE graphic arts composing equipment in 1963), and p. 378 (Medical Library Assistance Act of 1965).

Teletype. Vacuum tubes (cathode ray tubes, or CRTs, as they were called) and then, later, magnetic cores were the means for storage of operating programs and data during processing; but both were exceptionally expensive and limited in capacity (6,000 bytes would be a big internal memory). For large-scale data storage we had magnetic tapes and magnetic drums—each slow and with inherent limitations in the ways they could be used. For display we had the most primitive of CRT units, with low resolution and presenting only limited amounts of data. The means for output were punched-card tabulators, with upper-case-only fonts and which operated again at the equivalent of ten characters per second. Thus, even though the computers were fast, they could not do much, given the limited internal memory. They were inherently limited by capabilities for storage of large files and the slow speed of input and output.

The Publishing Context

During this period publishing was much as it had always been. Composition was essentially manual; although there were a few isolated experiments with computer technology, they were completely outside the mainstream of commercial publishing.

The form of publication was simply print, using the traditional means for achieving it. Xerographic means for duplication were just beginning to have an impact toward the end of the period. Indeed, in the late 1950s, when Walter Carlson was attempting to get companies interested in his invention, the response was almost universally "What's the market for replacing carbon copies?" Only the Haloid Corporation (which later became Xerox) in the mid-1950s was willing to take the leap into what became an overwhelming phenomenon.

The computer as a means for publishing simply had not yet arrived.

The Library Context

Where were libraries during this period? There was increasing recognition of the importance of libraries, both as part of the activities to which I have referred and independent of them. The Library of Congress had for many years been the primary center for production and distribution of catalog cards, through its Card Production Service. But that was primarily of significance to the libraries of the country, not to the using public.

What was significant to the public was the fact that Congress passed the Library Services Act (later expanded into the Library Services and Construction Act), which, among other things, fostered the creation of library networks that were of vital importance as automated systems became important to libraries (Holley, 1983). The Medical Library Act of 1956 made the Surgeon General's Library (first created in 1840) into the National Library of Medicine (Miles, 1982, pp. 353–355). In 1862 Congress had created the library of the Department of Agriculture, and in 1962 it too became a national library.[3] All of these acts were clear recognition of the importance of libraries.

Internal Technical Processing

The next thread in the set of timelines is the development of automation for internal technical processing in libraries. Until virtually the end of the period progress in this respect was essentially nil. Of course, there were the highly successful uses of microfilm in the management of circulation records, especially in the public libraries, and there were many abortive efforts to use punched-card equipment for that purpose. There were similarly abortive efforts to deal with serial records, again using punched-card equipment. But for the core technical service functions—acquisition and cataloging—there were not even any abortive efforts. The problems were too great, especially with respect to the number of catalog entries involved and the overwhelming costs in converting them to machine-processible form.

Among the abortive experiments, two were of historical interest. The first was the effort in 1930 by Ralph Parker at the University of Texas. Parker wanted to try using a punched-card system for circulation control. The library director, Don Coney (who later became Director of Libraries at the University of California, Berkeley) said, "OK. Here's $300, but use it wisely" (Hayes & Becker, 1970). The second was the effort, again by Ralph Parker, now at the University of Missouri, to initiate an evolutionary approach to an integrated library records system (Parker, 1952).

By the end of the period the Council on Library Resources had begun to play crucial roles in the successive stages in development of library automation. The council was especially important in providing support to the National Library of Medicine in its effort to auto-

[3] The Organic Act of 1862, which established the Department of Agriculture, clearly identified the need for a library within it, and the first librarian was appointed in 1867. Over the years, the Department of Agriculture Library became, de facto, a national library but it was not officially designated as the National Agricultural Library until 1962, when Memorandum No. 1496 of the Secretary of Agriculture did so.

mate the production of *Index Medicus* (Miles, 1982) and to the Library of Congress in its first explorations of the use of computer-based systems (King, 1963). More generally they were concerned with alternative means for producing catalogs and making catalog data available.

In that respect one of the important attempts to apply computers to provide access to catalog data was the production of book-form catalogs, especially for union catalogs, which contain records of a set of libraries. This approach has a history of some importance to the development of librarianship in the efforts of Charles Jewett to use stereotypes for the production of a national union catalog ("Fifth annual report," 1851, pp. 28–41, 81).[4] In Los Angeles the County Library system experimented with the use of punched-card equipment for that purpose (MacQuarrie, 1984), and there was a similar effort in Seattle, at the King County Library System. While those efforts were essentially dead ends in the context of later developments, they were important steps in the general progress.

Information Science and Information Services

Given the requirements for information storage and access, focused especially in the intelligence community, computers were seen as a potential means for meeting them. Despite the limitations during this period the expectations were that the technologies would steadily improve; so there were many efforts to solve the technical and theoretical problems in this field. Indeed, my own company, Advanced Information Systems, was established precisely for that purpose, and by 1964 we had developed the first commercially successful database management system using the technical skills generated through research projects for the National Science Foundation, the Air Force, and other agencies.

But that was toward the end of the period. At the beginning of it the methods for information retrieval were largely based on physical matching of search criteria with document data. In this vein Calvin Mooers had developed Zatocoding as a means for using edge-notched cards, and Mortimer Taube had developed implementations of his Uniterm concept (Taube & Wooster, 1958; Taube, 1959).[5] The Intellofax system at the CIA did much the same thing, using punched-card equipment, and Hans Peter Luhn at IBM developed a similar set of approaches (Schultz, 1969).[6] In a real sense this focus on physical matching reflected the very nature of punched-card logical processing, which was based on direct connection by wires on a plugboard.

At that time even the attempts to apply computer technology to the tasks in retrieval started with physical matching, using optical coincidence, of search criteria with document data. In particular, the Rapid Selector (the realization by Ralph Shaw, at the National Agricultural Library, of Vannevar Bush's Memex) used optical matching. The Minicard system, developed by Eastman Kodak and the Magnavox Company for the intelligence community, similarly used optical matching. Even the Western Reserve "Searching Selector," developed by James W. Perry and Allen Kent, used an electronic counterpart of such matching (Becker & Hayes, 1963). The point is that those early developments had not yet recognized the capabilities of the computer for complex processing of recorded symbols, so it was not until the end of this period that the techniques of modern computer-based retrieval began to appear. On the other hand, it must be said that during even the early stages of this period many of the tools for complex information processing began to be developed. In fact, experiments in automated language translation were already under way by 1950, and they continued to be a major research investment until the end of the period.

One development toward the end of this period, made operationally possible by use of computer technologies, deserves special recognition. That was the creation of the *Science Citation Index* by Eugene Garfield and his associates at the Institute for Scientific Information (ISI). It truly revolutionized the means for indexing the literature of science, placing it in the hands of its users (Garfield, 1977).[7]

---

[4] In *Fifth annual report*, see especially pp. 28–41 (Report of the Assistant Secretary in charge of the library) and p. 81 (Report of the Commissioners upon the general catalogue).

[5] Taube and Wooster contains articles by H. P. Luhn, Calvin Mooers, Ralph Shaw, and others, as well as by Taube himself. *Emerging Solutions for Mechanizing the Storage and Retrieval of Information* contains a description by Luhn of the "IBM universal card scanner" that involved optical coincidence as the means for selection. Contains a variety of articles concerning coordinate indexing written by Taube and his associates.

[6] In Schultz, see especially "The IBM electronic information searching system," pp. 35–51 and "Information retrieval through row-by-row scanning on the IBM 101 electronic statistical machine," pp. 164–185.

[7] In Garfield, see especially Weinstock, Melvin, "Citation indexes," pp. 188–195 (reprinted from *Encyclopedia of library and information science, vol. 5*, pp. 16–40. New York: Marcel Dekker, 1971.)

1964 to 1980

At the beginning of this second period I joined the faculty of the School of Library Service, at UCLA. I almost immediately became director of the University-wide Institute for Library Research, the special objective of which was to explore the issues involved in developing automation in libraries and especially in academic libraries, such as those of the University of California. During the middle of the period I took a brief leave from the university to return to the commercial sector. Joseph Becker and I founded Becker & Hayes, Inc., to provide consulting services in the development of library automation, with special emphasis on its use in library networks (such as that of Washington State). We were later acquired by the publisher John Wiley and Sons and for some time published a series of books in the field of automation and information science.

The Scientific, Social, Economic, and Political Context

In retrospect, although it was not evident at the time, there was a steady reduction in the cold war conflict between the United States and the Soviet Union. The tensions were still there, and the cold war conflict would periodically bubble up and then simmer down. But the focus moved on to Vietnam and a real war, one that was insanity on our part and that may have virtually destroyed our society. As I recall, the cold war tensions were no longer the driving force for developments in our field.

In fact, something far more fundamental and ultimately revolutionary had become the driving force. That was the transition of the United States and, at a much slower pace, other industrialized countries from being "industrial" economies to being "information" economies. Today we see the impact of that transition in its effect on every component of our economy and our society, but it was during this period that the changes became evident. The report by Marc Porat on the information economy of the United States clearly identified what was happening and the fact that by the mid–1970s over 50 percent of the nation's workforce was engaged in information work (Porat, 1977).

Although many equate the appearance of the information economy as something created and driven by information technology, in my view the cause is more fundamental. It is the imperative in societal development that has created what we see, and the technology merely feeds and serves that imperative, making it possible but not causing it. In any event it is the fact of the information economy, whatever the cause, that was the driving force for developments in our field during this period. These perceptions led to the belief by some that there should be formal recognition of the need for a "national information policy" in the United States ("National information policy," 1976). Several other countries had already begun to develop national information policies, not least among them Japan, which made this a major priority for its Ministry of International Trade and Industry.

The importance of science information continued to be recognized. The Department of Commerce established the National Technical Information Service (NTIS) as one means to improve dissemination of scientific and technical information. The National Academy of Sciences repeatedly reviewed the status of developments and recommended increased efforts ("Science information activities," 1965; Committee on Scientific and Technical Communication, 1969).

From the standpoint of scientific information, the dominance of the United States, so evident during the previous period, began to dissipate. Several countries established scientific and technical information centers to coordinate national access to worldwide scientific information. I think especially of the Japanese Information Center for Science and Technology (JICST), in Japan, but there were similar developments in Taiwan, Saudi Arabia, and elsewhere (Hayes, 1972). By the end of the period the United States produced less than 40 percent of the world's scientific publication, and the rate of decline appeared to be about 3 percent per decade (National Science Board, 1989, p. 327).

Commercial, Industrial, and Consumer Context

During this period the importance of information began to be recognized by commercial and industrial companies. They began to install centralized management information systems and to experiment with access to external databases. While the level of information use was still small in comparison with today, it was clearly growing.

Among the developments during this period were several that paralleled those in information science. They were systems that provided online access to a variety of commercial databases, such as those for checking the credit status of individuals and organizations ("Survey of credit card verification systems," 1971), those for airline reservations management,[8] and those for stock market

[8] The Saber system, developed by American Airlines and still operative today, comes to mind.

quotations. The Quotron system was an early example of the latter, and it is still operating today on the Internet. The technologies involved, combining as they did computers with telecommunications, were essentially the same as those for reference database access services. The software, though, was different because it involved much simpler record structures and simpler retrieval criteria, but far greater numbers of transactions, by orders of magnitude.

The consumer use of information was also one of growth, especially for the various entertainment media—movies, network and cable TV, and sports. From 1964 through 1980 expenditures for consumer information quadrupled, increasing from $20 billion to $80 billion per annum. Of that, sales of leisure books, newspapers, and magazines also quadrupled, from about $ 10 billion to $40 billion.

Information Technology

During most of this period computing was centered on large mainframe computers. In fact, what was called "Grosch's law" (named after Herbert R. J. Grosch, who formulated it) governed most of the decisions about computing installation (Orr, 1968, p. 152). It states that "computing power goes up as the square of the cost," the implication of which is crystal clear: The bigger the better! During this period the California state legislature became committed to that view and attempted to force the University of California to use a single computing center to serve all nine campuses. The MELVYL system (named in honor of Melvyl Dewey of the Dewey Decimal System) was initially visualized as a centralized university-wide online catalog in the spirit of Grosch's law.

In parallel, there was extensive development of networks among computers, starting from the work of the UCLA Western Data Processing Center (the first such network) and expanding into ARPANET, of the Department of Defense, and the NSF's supercomputer network. Together they became the backbone of today's Internet (now supplemented by all of the commercial communication networks). These computer networks were made operationally feasible by implementing satellite and fiber-optic communication systems and by developing packet switching to break messages into pieces that could be sent independently by the fastest paths through the network and then reassembled at their destinations.

Publishing

During this period the processes of publishing were revolutionized, as computer-based photocomposition completely replaced the former manual methods. Xerography became a fact of daily life in business of every kind. From the standpoint of this history, though, the important development in publishing was the onset of electronic formats. The beginning was the creation of databases, initially as a result of efforts by the U.S. federal government.

The Library Context

During this period efforts increased to establish cooperative networks among libraries—not so much in the form of communication networks as in the form of administrative networks. Each of the states, under the stimulus of the Library Services and Construction Act, created its own multi-type state library network. The National Library of Medicine, under the mandate of the Medical Library Services Act, created the regional medical library system as a network. A variety of other networks were formed among groups of other types of libraries, especially academic ones. I recall especially the "Harvard-Yale-Columbia Medical Libraries" partnership that gave Fred Kilgour the springboard from which he created the Online Computer Library Center (OCLC) (Kilgour, 1984).[9] Indeed, this effort at cooperation reflected the generic issue of retrospective conversion of catalog records, which became a dominant concern of the academic library community. It was the reason for creating both OCLC and the Research Library Network as competing answers.

Cooperation in the area of cataloging was made possible only because of the crucial contribution of Henriette Avram at the Library of Congress in establishing the MARC (machine readable cataloging) format as the de facto national standard for exchange of catalog data. Without that it would have been intolerably difficult (Avram, Knapp & Rather, 1968; Avram, 1968; Avram, 1975).

Internal Technical Processing

It is in this area that perhaps the most dramatic change took place within this period. With the stimulus of the MARC format and the use of OCLC and the Research Libraries Information Network (RLIN) as economic solutions to the catalog conversion task, systems to

[9] In Kilgour, see especially pp. 235–238, 265–270, and 309–314.

support cataloging could be developed within institutions. At least a couple of them—DOBIS, developed at Dortmund University in Germany, and NOTIS, developed at Northwestern University in Illinois—became commercial products.

In parallel, modules were developed to provide other aspects of internal technical services—circulation and collection management, serial records, and acquisitions.

Information Sciences and Services

The major activities during this period are so well presented in the *Annual Review of Information Science and Technology* that I am not going to do more than highlight what to me were some of the most significant of them (Cuadra, 1966–present).[10] First, as a result of all that had occurred during the first period, the production and distribution of databases became a reality rather than merely a speculation. All of the research and technical development had come to fruition, and with great success. Of special importance were the online services (DIALOG, for example) for access to the reference databases (such as Chemical Abstract Services, Educational Resources Information Center [ERIC] , and MEDLINE) and library OPACs (online public access catalogs) for access to catalog databases. Less dramatic but still important was the implementation of interlibrary loan (ILL) services by OCLC and RLIN.

Second, beyond these operational developments were the academic ones, as information science became part of university education. At the end of the previous period there was discussion of the need for educational programs in science information and information science (Crosland, 1962; Goldwyn & Rees, 1965). At the beginning of this period persons who had been in industry joined library schools and introduced information science into the professional curriculum. For example, I joined the School of Library Service at UCLA, Don Swanson that at Chicago, Allen Kent and James Perry that at Western Reserve, and later Allen Kent that at Pittsburgh. At the same time programs in computer science were established, most in engineering schools but a few in departments of mathematics, and many of them incorporated elements of information science into their curricula. Schools of management later implemented programs in management information systems, again many of which also incorporated information science.

At the time none of these educational programs was unequivocally accepted as academically legitimate. I can recall the debates in 1966 when I served as chair of the Academic Senate committee that assessed whether a computer science department should be established at UCLA. The initial view of some on the committee was that computer science really was not an academic discipline worthy of separate recognition. Fortunately, the case was well made, and the decision was unanimously in favor. But it did require argument.

Third, at least at the beginning of this period, there were many discussions and in some cases explorations of alternative ways of providing science information services, related to and to varying extents dependent on the methods being developed in information science, but involving rather different kinds of approaches. Among them were recommendations that a national agency should be established to be responsible for coordination of science information, even perhaps modeled on the Soviet system embodied in VINITI (All-Soviet Institute for Scientific and Technical Information); fortunately, the pluralistic approach that characterizes the United States prevailed (National Information Center, 1963).

Others involved efforts to build upon, support, and use the “invisible colleges.” Those efforts became quite controversial at the time and were abandoned, as such. But today they have been realized through the Internet in the various “list-servs” that support such direct and immediate peer-to-peer communication. Related approaches involved discipline-specific efforts like those of the Biological Science Communication Project (Janaske, 1962; Hattery, 1961). Perhaps the most significant and long-lasting of them were the several discipline-specific information centers that served as focal points for review and analysis of literature relevant to current research (“Management of information analysis centers,” 1972). I think especially of the ones established by the National Institutes of Health, among them the Brain Information Service at UCLA; of the National Standard Reference Data Center at the then–National Bureau of Standards (Brady, 1968; Rossmassler, 1972); and of the several ERIC Clearinghouses (Burchinell, 1967).

Repeatedly, from the beginning of the 1960s to the present, congressional committees have reviewed the status of all of these experiments, and the related reports are excellent sources for perspective on the relevant history (“Documentation, indexing, and retrieval,” 1960; “Documentation, indexing, and retrieval,” 1961;

[10] The successive volumes of the Cuadra review cover not only the essential developments but provide a picture of the changing emphases in research and development in the field each year.

"Interagency coordination of information," 1962; "Scientific and technical information (STI) activities," 1978; "Information and telecommunications," 1981).

1980 to 1998

The Scientific, Social, Economic, and Political Context

This has been a most remarkable eighteen years! We saw the end of the cold war with the collapse of the Soviet empire, which led to the disintegration of other federations—Yugoslavia and Czechoslovakia, in particular. Those two for me were special shocks, since I had developed close ties with each of them.

The effects of these international events for our field are less tangible perhaps, but they are very real. The separate republics of the former Soviet Union became independent countries, each of which will need to deal with the information revolution. Beyond them other countries of central and eastern Europe that had functioned under the stultifying effects of Soviet-style communism face similar needs.

In the United States the continued progress in development of the Internet led the Clinton administration to identify the need for a national and then international "global information infrastructure." For the first time "information" beyond simply science information had become an explicitly identified priority in national policy (Clinton & Gore, 1992; "NTIA infrastructure report," 1991).

Commercial, Industrial, and Consumer Context

This period has seen a virtual revolution in the extent to which information resources are used throughout the United States. Today massive amounts of data are generated, transmitted, and consumed. The cellular telephone is ubiquitous, even in central and eastern Europe, where it serves as a substitute for the lack of adequate telecommunication infrastructure for the nongovernmental economy. The Internet and World Wide Web are growing in use at a phenomenal rate—doubling every three to six months. The entertainment and amusement industries are exploding. The publishing of books and magazines has grown similarly. Indeed, the "super bookstores" proliferate at an almost unbelievable rate, and one of the success stories of the Internet is amazon.com, an online bookstore.

Within companies, management information systems now function as decentralized services that bring data directly to the point of immediate need. Communication within companies and with their customers invariably starts with bringing up a display of an appropriate record from a file. The use of the Internet and its services brings external data directly into the process of decision making. For the consumer the picture is comparable. Today online banking from the home is a fact of life, and the use of credit cards in every commercial venue is commonplace.

Information Technology

You will recall that during the first thirty years or so of this history the decisions concerning computer acquisition were based largely on Grosch's law: Bigger is better! But something happened toward the end of the 1980s that was a fundamental revolution. The microcomputer—the PC and the Macintosh—totally reversed the law: Smaller is better! The new law is "Moore's law" (named after Gordon Moore, cofounder of Intel), which states that the capabilities of microprocessors double every two years (Moore, 1965). The result is that today, as I said before, I can hold in the palm of my hand more computing power than even the largest of mainframes had two decades ago. The result has been a distribution of computing power that puts the PC, laptop, or palmtop in the plane seat, the police car, the fire engine, the personal auto, as well as in virtually every home and office.

The revolution in information technology is far greater than just the computer itself, though the computer is indeed the centerpiece. Increases in telecommunication capacity in some respects equal those in computer capacity. Whereas fifty years ago we were limited to ten characters per second (roughly the equivalent of 100 baud), today we have data rates five hundred times greater for use in our own homes. Whereas forty years ago we had CRTs with minimal capability, today we have SVGA displays with resolutions virtually the equivalent of the printed page and screen capacities that permit the most beautiful images imaginable and the ability to observe in real time the operation of the heart or the brain of a human being, in living color. Whereas fifty years ago we were happy to have 6,000 bytes of internal memory, today we can have almost unlimited numbers of megabytes. And whereas thirty years ago the means for mass storage of data were limited in capacity and unbelievably slow, today we have gigabytes of capacity with rapid random access.

What a revolution it has been!

Publishing

During this period publishing has continued essentially unchanged, though with steadily increasing concern about how best to deal with the impact of electronic formats.

The Library Context

Turning to the library context in the United States, the major effect of our virtual bankruptcy resulting from the huge commitment of resources to military buildup has been the economic pressures on any and every public enterprise, among which are libraries. The funding problems for both public and academic libraries have been real and in some cases catastrophic. University libraries experienced 10 to 20 percent budget reductions each year for five years and more. The price of materials skyrocketed; journals in particular escalated in cost at 15 to 20 percent a year over the past decade.

Yet, in the face of the economic travails, the use of libraries has grown dramatically, for both public and academic libraries. The information economy requires the kinds of resources and means for information access that libraries and only libraries provide.

Internal Technical Processing

During this period the bits and pieces that were created by individual institutions have been replaced by integrated library systems, commercially available and with the support of highly qualified professional staffs. They are operational on both mainframes and personal computers; they function in both stand-alone and "client-server" modes; they will serve every type and size of library; they will function well in any country of the world.

When I think back to thirty years ago, I recall an article written by Ellsworth Mason. Titled "The Great Gas-Bubble Prick't," it treated the use of computers in libraries as a con game of the computer enthusiasts (Mason, 1971). During the weeks after it was published, my colleagues on the faculty of the library school would gleefully mutter, "The emperor's clothes, hee, hee, hee!" as they passed me in the hall. The interesting thing is that, in my correspondence with Mason before he wrote that article, I cautioned him about the costs involved in automation and about the difficulties and uncertainties (Hayes & Mason, 1971). But I also stressed the values that it would have in the increased services available to the users. That is the reason for libraries and that is what automation provides. All else is but froth.

Information Sciences and Services

The effects on information science and information services have been dramatic. The availability of OPACs in virtually every library makes the resources of the library and, in most cases, of the world readily available. The availability of CD-ROMs brings a wealth of materials into the library, not only available but in processible form. The Internet and the World Wide Web provide means for online communication and access that are changing the entire information economy.

1999 to 2005

It is difficult to make predictions, especially about the future, but I am going to do so by adding a fourth period—the coming five years rather than more—to those I have already discussed.

The Social, Economic and Political Context

I predict that the revolution that has resulted in the "information economy" will continue and that the pace will accelerate rather than decelerate, at least in other countries if not in the United States. The fact is information products and services have become a worldwide phenomenon. It has been estimated to be a trillion-dollar market, and the United States is the dominant supplier. The Internet and World Wide Web are simply a manifestation of it. The entertainment industry is without doubt the major component, and computer-related software and information packages are of increasing importance.

The reason I predict that this phenomenon will continue to increase is both simple and complex. The simple fact is that it has become easy to create information-based products and services. As the basis for doing so grows, the capabilities increase exponentially, and new products provide the basis for newer ones. The more complex fact is that creating information products and services consumes almost no physical resources; this means that it is not only easy but also highly economical to do so. One can create a product or service with almost no investment, except one's time, and easily test the market for it at minimal risk.

Of course, to do so depends upon both technological infrastructures—the telecommunications systems and the availability of inexpensive computers—and the capability to use them. That means that societies and countries without that infrastructure and those skills will be less able to participate in the growth. The gap between the "haves" and the "have-nots" will grow, not decrease.

Furthermore, none of the information technologies in themselves will solve the fundamental problems of overpopulation or the irrationalities of intolerance and hatred. Much though I wish it were otherwise, we will continue to see all of the evils that have characterized inhumanity.

Commerce, Industry, and Consumers

So what is going to happen to the use of information in commerce, in industry, and by the consumer? Frankly, awesome though the implications are, I think it will continue to grow as it has for the past period.

Information Technology

During the fifty years that I have been in the computer business, it has continued to grow at a phenomenal rate. Each year the capabilities have increased, the effectiveness and even efficiency have increased, and the pervasiveness in our society has increased. Of that fifty-year history, though, the most spectacular gains have been made within the past twenty-five when microprocessors—computers on chips—became the basis for the hardware. The first microprocessors were breathtaking innovations, and they revolutionized not only the design of computers but also their use throughout the world.

Will the process continue? In a recent book Michael Malone poses that question and concludes that the answer is "yes," saying that while there are absolute barriers in the laws of physics, there are still means by which the process can continue (Malone, 1998). Most important among them are those that relate to the software, in which future developments may have even greater impact than we have seen to date from the hardware.

Publishing

What is the future of publishing, especially in light of what is happening with the information technologies? In particular, what will happen to publication in print formats? Let us look at the means for electronic publication.

First among them is the Internet and the World Wide Web. The Web in particular is a phenomenon that has almost literally exploded within the past four to five years. Indeed, Web use is growing exponentially, doubling every three to six months. That pace in growth is likely to continue for the next two years, though with the likelihood of significant slowing after that. Future increases in these services and uses, however, will be highly commercial, with advertising playing an increasing role and such uses as pornography being among the most evident (as they were with the French MINITEL service). The Internet, highly academic in its origins and orientation, while growing, is doing so at a substantially more limited rate and is already very small in comparison to the Web. Access to the Internet will continue to be important for its essentially academic roles (such as scholarly e-mail and search of library catalogs). For library planning it will be essential to control carefully the extent to which access to the Web consumes library resources, especially of time at terminals.

Currently, CD-ROM is the fastest growing form of publication, doubling in sales every two years. While it may soon be displaced by DVD (digital video disk), the trends will continue and be similar. A similar rate of increase is likely to continue for the coming ten-year period, though with a steady leveling over time. In the past publications of importance to libraries have been of reference materials (indexes, abstracts, encyclopedias, data collections, etc.), which typically are made available through the CD-ROM local area network; those publications are likely to level off quite rapidly. The substantial increases will be in instructional materials, especially as represented by multimedia packages; in journals, especially scientific journals; and perhaps in other types of material. For purposes of general reading, however, this form of electronic publication is unlikely to displace the printed book or popular journal. Therefore, for library planning, it is necessary to plan for continued acquisition of materials in print form.

Electronic document delivery is a reality today, as represented by the commercial services that provide it as a replacement for interlibrary lending as well as for its own values in speed of delivery. It is also represented by a few existing journals. For example, some newspapers are already distributing Web versions of their materials. Some popular journals are likely to begin to appear through the Web, if they have not done so already. OCLC is publishing several scholarly journals online, such as the *Journal of Current Clinical Trials.* A variety of online communications on the Internet effectively serve as journals for limited groups. In the future general scholarly journals, especially those produced by commercial publishers, are likely to shift from current print publication of issues to future on-demand publication of articles. Print versions may even disappear. Other serial publications, such as conference proceedings, are also likely to be distributed in online form (as well as in CD-ROM format). There are no data for projecting the pace at which this will happen, however.

While there is clear evidence for a shift in journal publication from printed issues to on-demand publication of articles, there is no evidence of comparable shift of monograph publication in the same way. And even though newspapers and popular journals are likely to publish online versions, distributed through the Web, they will continue to publish their print versions. As with CD-ROM materials, therefore, for purposes of general reading, online electronic publication is unlikely to displace the printed book or popular journal. Again, for library planning, it is necessary to plan for continued acquisition of materials in print form.[11]

Digitized imaging is a form of publication that will be of increasing importance in specific areas. It has become a powerful tool in research of all kinds. It is the basis for publishing of such materials as maps and may well replace print. Academic libraries in particular should plan for management of digitized images.

Libraries

What will happen to libraries? There are persons who forecast their demise in the perception that they will be replaced by the wealth of resources becoming available through the information technologies; such voices have been heard for at least the past three to four decades. It is a fact that during the past decade libraries have faced enormous economic pressures; they have had to operate within the constraints of reduced budgets at the same time that the costs of acquisitions, especially of journals, have been escalating. At the same time they have needed to make continuing investments in automated systems and to deal with the array of computer-based forms of publication.

My own perception, though, is very different from that of those who wish to get rid of libraries. My view is that libraries are essential and will continue to be so in the foreseeable future. Instead of being overwhelmed by technologies, they have absorbed them, made them economic and effective, and served as the basis for testing and proving them. It is also a fact that the use of libraries of every type has been increasing, not decreasing. Indeed, a study at Columbia University showed that the effect of electronic information resources was to increase not decrease the use of the library. The various forms of publication are complementary and mutually supportive rather than being substitutes for each other. The use of any of them leads to increased use of the others, and the library serves as the agency for access to all of them.

I have every expectation that the library will continue not only to exist but to thrive and to play its historic leadership role in the coming decades. Underlying that expectation is my view that, while electronic publication will be increasingly important, it will not replace print in the foreseeable future. And libraries will be the means of access to both print and electronic formats.

Library Automation

My perception of internal technical processing is that the dramatic changes resulting from automation have already occurred and that the future will not significantly add to them. The bibliographic utilities, OCLC especially, are well-established, economically viable operations; their scope of coverage is becoming increasingly international. Cataloging and acquisitions work can depend upon their online union catalogs (OLUC, as OCLC refers to it).

Having said that, it is likely that there will be a steady shift to outsourcing of cataloging and even of acquisitions work, at least within smaller academic and public libraries. (For the special libraries that serve business and industry, it is already a fact of life, and they really function primarily as a means for access to information rather than as collections of materials.)

The implications are that library services to patrons will become increasingly important and that staff will be shifted from internal operations to direct services.

Information Sciences and Services

What will happen to information science and information services? The crucial point to me is that the widening scope of information resources increases the importance of both the science and the services. In this respect we should recognize that the library is more than simply a collection of materials, valuable though that is and will continue to be. It is also more than simply a means for access to those materials and the information contained in them, again valuable though that is and will continue to be. The library is the agency that serves as the means for selection. It does so when acquisition librarians make decisions about what materials are worth adding to the collection. It does so when reference librarians help patrons in locating and selecting from the wealth of resources those that will meet needs. It does so when library-based information specialists select from retrieved information and analyze the results.

The library is also a means for users to learn how to manage information resources. Library services in teaching are therefore of special significance, and the increasing wealth of resources adds greatly to their importance.

[11] See, for example, Textbook case: publishers look to cash in on rising demand for books, *Barron's*, 7 Dec 98: “Experts predict that elementary and high school textbook sales will jump by 35%, to $3.5 billion, from 1996 through 2001. College textbook sales are expected to zoom 40%, to $3.47 billion, over the same span.” (http://www.smithbarney.com/cgi-bin/bench)

References

Avram, H. D. (1968). *The MARC pilot project: Final report on a project sponsored by the Council on Library Resources.* Washington, DC: Library of Congress.

Avram, H. D. (1975). *MARC, its history and implications.* Washington, DC: Library of Congress.

Avram, H. D., Knapp, J. F., & Rather, L. J. (1968). *The MARC II format: A communications format for bibliographic data.* Washington, DC: Library of Congress, Information Systems Office.

Becker, J., & Hayes, R. M. (1963). *Information storage and retrieval.* New York: Wiley.

Brady, E. L. (1968). *National Standard Reference Data System.* Washington, DC: National Bureau of Standards.

Burchinell, L. G. (1967). *ERIC and the need to know.* Washington, DC: Government Printing Office.

Bush, V. (1945a, July). As we may think. *Atlantic Monthly, 176,* 101–108.

Bush, V. (1945b). *Science: The endless frontier; Report to the President on a program for postwar scientific research.* Washington, DC: Government Printing Office.

Clinton, W. J., & Gore, A. (1992). *Putting people first: How we can all change America.* New York: Times Books.

Committee on Scientific and Technical Communication, National Academy of Sciences, National Research Council. (1969). *Scientific and technical communication: A pressing national problem and recommendation for its solution (The SATCOM report).* Washington, DC: National Academy of Sciences.

Crosland, D. M. (1962). *Proceedings of the conferences on training science information specialists: October 12–13, 1961, and April 12–13, 1962.* Atlanta: Georgia Institute of Technology.

Cuadra, C. (Ed.). (1966–present). *Annual review of information science and technology.* Washington, DC: American Society for Information Science.

Current research and development in scientific documentation. (1957–1969). No. 1–15. Washington, DC: National Science Foundation, Office of Science Information Service.

Documentation, indexing, and retrieval of scientific information; Report of the Committee on Government Operations, United States Senate. (1960, May 24). Washington, DC: Government Printing Office.

Documentation, indexing, and retrieval of scientific information: A study of federal and nonfederal science information processing and retrieval programs. Committee on Government Operations, United States Senate. (1961, March 9). Washington, DC: Government Printing Office.

Fifth annual report of the Board of Regents of the Smithsonian Institution. (1851). Washington, DC: Government Printing Office.

Garfield, E. (1977). *Essays of an information scientist.* Philadelphia: ISI Press.

Goldwyn, A. J., & Rees, A. M. (Eds.). (1965). *The education of science information personnel—1964.* Cleveland: Western Reserve University, School of Library Science, Center for Documentation and Communication Research.

Hattery, L. H. (1961). *Information and communication in biological science: A report prepared for the Biological Sciences Communication Project, American Institute of Biological Sciences.* Washington, DC: American University Center for Technology and Administration.

Hayes, R. M. (1972). Science information in the Republic of China. *Bulletin of the National Science Council of the Republic of China.* Taipei: National Science Council, Republic of China.

Hayes, R. M. (1992). The Defense Logistics Agency and the Federal Catalog System. *Government Information Quarterly, 9,* 291–303.

Hayes, R. M., & Becker, J. (1970). *Handbook of data processing for libraries.* New York: Wiley, pp. 3–4.

Hayes and Mason on automation . . . (1971). [Exchange of correspondence]. *College and Research Libraries,* 32, 388–392.

Holley, E. G. (1983). *The Library Services and Construction Act: An historical overview from the viewpoint of major participants.* Greenwich, CT: JAI Press.

Information and telecommunications: An overview of issues, technologies, and applications. (1981, July). Report for the Subcommittee on Science, Research and Technology of the Committee on Science and Technology, U.S. House of Representatives. Washington, DC: Government Printing Office.

Interagency coordination of information: Hearings of the Subcommittee on Reorganization and International Organization of the Committee on Government Operations, United States Senate. (1962, September 21). Washington, DC: Government Printing Office.

Janaske, P. C. (Ed.). (1962). *Information handling and science information: A selected bibliography, 1957–1961.* Prepared by the American Institute of Biological Sciences, Biological Sciences Communication Project in cooperation with the American University Center of Technology and Administration, School of Government and Public Administration. Washington, DC.

Kilgour, F. G. *Collected papers of Frederick G. Kilgour, early years.* (1984). P. A. Becker & A. T. Dodson (Comps.); L. L. Yoakam (Ed.). Dublin, OH: OCLC Online Computer Library Center.

King, G. W. (1963). *Automation and the Library of Congress.* Washington, DC: Library of Congress.

MacQuarrie, C. (1984). The metamorphosis of book catalogs. *Library Resources and Technical Services,* 8, 370–378.

Malone, M. S. (1998). *The microprocessor: A biography.* Santa Clara, CA: Telos.

The management of information analysis centers: Proceedings of a forum sponsored by the COSATI Panel on Information Analysis Centers. (1972). Washington, DC: Committee on Scientific and Technical Information, Federal Council on Science and Technology.

Mason, E. (1971). The great gas bubble prick't; or, Computer revealed—by a gentleman of quality. *College and Research Libraries, 32,* 183–195.

Miles, W. D. (1982). *A history of the National Library of Medicine.* Washington, DC: Government Printing Office.

Moore, G. E. (1965, April 19). Cramming more components into integrated circuits. *Electronics,* 114ff.

National Information Center. (1963). Hearings before the ad hoc Subcommittee on a National Research Data Processing and Information Retrieval Center, held in Washington, DC, 27–28 May, 18–19 July, and 17 and 19 September.

National information policy: Report to the President of the United States submitted by the staff of the Domestic Council Committee on the Right of Privacy. (1976). Washington, DC: National Commission on Libraries and Information Science.

National Science Board. (1989). *Science and engineering indicators—1989.* Washington, DC: Government Printing Office. (NBS 89-1) .

Nonconventional scientific and technical information systems in current use. (1958–1966). No. 1, 1958; No. 2, 1960; No. 3, 1962; No. 4, December 1966. Washington, DC: Government Printing Office.

The NTIA infrastructure report: Telecommunications in the age of information. (1991). Washington, DC: U.S. Department of Commerce, NTIA.

Orr, W. D. (1968). *Conversational computers.* New York: Wiley.

Parker, R. (1952). *Library applications of punched cards.* Chicago: American Library Association.

Porat, M. U. (1977). *The information economy. Definition and measurement.* Washington, DC: U.S. Department of Commerce, Office of Telecommunications.

Rossmassler, Stephen A. (Ed.). (1972). *Critical evaluation of data in the physical sciences—a status report on the National Standard Reference Data System.* Washington, DC: National Bureau of Standards.

Schultz, C. (Ed.). (1969). *H. P. Luhn: Pioneer of information science, Selected works.* New York: Spartan Books.

Science information activities: A report of the Office of Documentation. (1965). Washington, DC: National Academy of Sciences, National Research Council.

Scientific and technical information (STI) activities: Issues and opportunities. (1978, December). Prepared for the Subcommittee on Science, Research and Technology of the Committee on Science and Technology, U.S. House of Representatives. Washington, DC: Government Printing Office.

Specialized science information services in the United States. (1961, November). Washington, DC: Government Printing Office.

Survey of credit card verification systems. (1971, March 29). *Electronics,* 45.

Taube, M. (Ed.). (1959). *Emerging solutions for mechanizing the storage and retrieval of information* (Studies in Coordinate Indexing, Vol. 5). Washington, DC: Documentation, Inc.

Taube, M., & Wooster, H. (Eds.). (1958). *Information storage and retrieval: Theory, systems, and devices (Air Force Office of Scientific Research Symposium).* New York: Columbia University Press.

On the Shoulders of Giants

Eugene Garfield

The title for this talk was inspired by Robert K. Merton's *On the Shoulders of Giants* (1993). This book is an unusual exercise in scholarly detective work, which is often referred to as *OTSOG*. It tracks down the origins of Isaac Newton's famous aphorism, "If I have seen farther, it is by standing on the shoulders of giants"—an observation that certainly applies to my career in information science and science communication. I would like to remember tonight many giants whom I have known and whose contributions and personalities made my career a very enjoyable one. It has been a privilege to be associated with such people.

Early Contacts

Robert K. Merton should not be confused with his Nobel Prize–winning son, the Harvard economist Robert C. Merton. Robert K.'s association with the field of information science is quite strong. He has probably been cited by information scientists as often, if not more so, than many of those we usually recognize as pioneers of information science. This is partly because of his pathbreaking work in the sociology and history of science, but also because he is the inventor of numerous neologisms now in common use in the field of information science. His "Matthew effect" (Merton, 1968a; Garfield, 1982b) has been cited in hundreds of papers and is itself the subject of numerous research papers, including those by my dear friend Manfred Bonitz of Dresden in Germany (Bonitz, 1997). The Matthew effect is manifest, for example, in the unfair attribution to a single senior author of work by two or more authors. Bob's wife, Harriet Zuckerman, has sometimes been a victim of this effect. She is well known for her work on Nobel Prizes (Zuckerman, 1996), but the papers she and Bob coauthored are often cited as "Merton and Zuckerman" when in fact the by-line was "Zuckerman and Merton." Another one of Merton's terms is OBI—"obliteration by incorporation" (Merton, 1968b, pp. 27–29, 35–38; Garfield, 1975), and all citation analysts will attest to the frequency of this phenomenon.

In 1962, the year that I met Bob Merton and Harriet Zuckerman, I wrote an almost identical letter to Bob, Derek De Solla Price, and J. D. Bernal suggesting that they might be interested in the Institute for Scientific Information's first experimental citation printouts.

British-born and -educated Derek Price was a physicist, historian of science, and an energizer of scientometrics worldwide.

Derek's countryman and physicist Bernal was the acknowledged initiator of the field of "Science of Science," which was a precursor of the social studies of science and the field of scientometrics. See my essay about the Bernal Award of the 4S Society established by ISI in 1981 (Garfield, 1982a). My first "contact" with Bernal was even before World War II, when my uncle presented me with a copy of Bernal's 1939 book, the *Social Function of Science*. I was just fourteen at the time. My first *professional* contact with Bernal occurred in 1958, shortly before the International Conference on Scientific Information in Washington, D.C. (*Proceedings of the International Conference on Scientific Information,* 1959).

The term "scientometrics" was coined by our beloved Russian colleague Vassily V. Nalimov. We first met at the Moscow Book Fair. Four of his books were published by ISI Press in translation.

One of Bernal's many disciples was the British physicist and journalist Maurice Goldsmith, who wrote a biography of Bernal (1980), and just before he died, a biography of Joseph Needham (1995). Maurice was very close to Federico Mayor, director general of UNESCO,

Robert Merton, Harriet Zuckerman, and Eugene Garfield.

Eugene Garfield and J. D. Bernal.

Eugene Garfield and Derek De Solla Price.

Eugene Garfield and Vassily Nalimov at the Moscow Book Fair.

Calvin Lee, Federico Mayor, and Eugene Garfield.

which published the Needham biography. I first met Mayor in 1965 when he was rector of the University of Granada, where he invited me to lecture about the *Science Citation Index* and *Current Contents.* Calvin Lee and I met with Federico in Paris on one occasion. Lee was my personal assistant at ISI for over a decade. He was an information scientist in London when we first met at OSTI. He is retired. He also accompanied me and Dr. Sher to China long before it was fashionable.

Another pioneer of information science, not listed as an ASIS pioneer, is Joshua Lederberg. I met him in 1959 shortly after he had won the Nobel Prize in medicine.

At a critical juncture he weighed in with his support for the concept of citation indexing and encouraged me to seek support from the National Science Foundation to conduct a pilot project [see http://www.profiles.nlm.nih.gov/BB/]. Josh still occasionally sends handwritten notes like his note of 9 May 1959, but he was also a pioneer in using e-mail. His Eugram was an early predictor of e-mail, the Internet, and electronic journals (Lederberg, 1978). Josh eventually became a member of the ISI board of directors. Our correspondence from 1959 to 1962 is already posted on my Web site (http://165.12.33.33/eugene_garfield/lederberg/list.html.

But Josh, J. D. Bernal, Derek Price, and Bob Merton were not my first contacts in the field of science information. In early 1951, after an accidental explosion at Columbia University in Louis P. Hammett's laboratory, I decided to look for a new job as a chemist. So I attended the spring seventy-fifth anniversary meeting of the American Chemical Society in March 1951, where I met James W. Perry, then at MIT, who was chairman of the Division of Chemical Literature. After the meeting I went up to him and asked, "How do you get a job in this racket?" I realized from attending that meeting that people were getting paid for doing something I gladly did for nothing. About a month later, after dining on several of my mother's wonderful Jewish meals, he introduced me to Sanford V. Larkey, the director of the Welch Medical Library at Johns Hopkins University. (Through Jim Perry I also met his longtime associates Madeline Berry Henderson as well as Allen Kent.)

The Welch Project

The Welch Project was established to find solutions to the Army Medical Library's retrieval problems and to evaluate machine methods for indexing. Every day San Larkey and I would work across the table from each other, and then we would go out to lunch at a local bar. He would tell me all about his army experiences and his interests in Elizabethan medicine, and then we would discuss subject headings. We had a fight once about whether "socialized medicine" would ever be accepted as a MeSH heading. Eventually, he convinced Brad Rogers to use "medicine, social"—or something like that.

The Welch Project afforded me the unique opportunity of meeting most of the ASIS pioneers. The Committee of Honorary Consultants to the Welch Project included, among others, Mortimer Taube, Verner C. Clapp, and Ralph Shaw. We are indebted to Mortimer Taube for coordinate indexing and much else.

Peter Luhn and Herbert Ohlman simultaneously but independently invented the now familiar key-word-in context (KWIC) indexing. In 1951 Pete came to see how we used the IBM 101 statistical machine at the project.

Verner Clapp was a great gadgeteer. As president of the Council on Library Resources, he later funded the development of the Copywriter—a device that I developed for selectively copying references (Garfield, 1973).

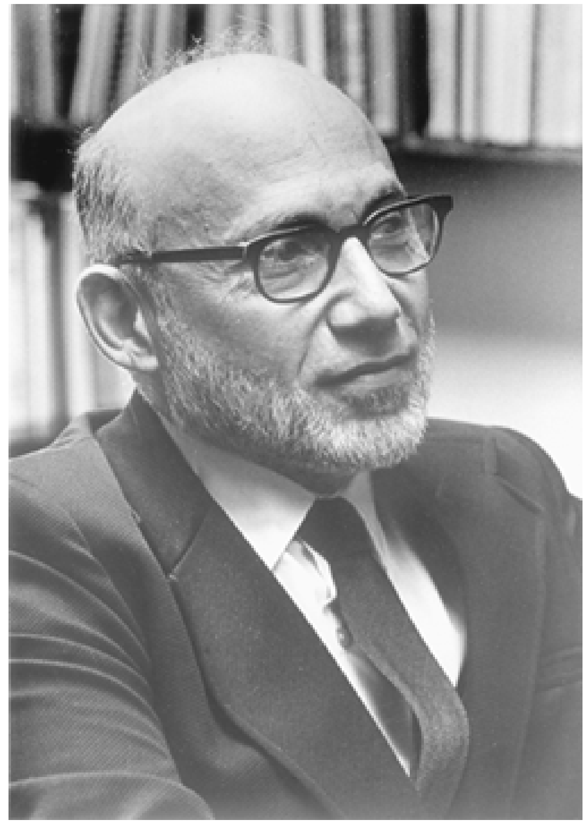
Ralph Shaw (Garfield, 1978b), then director of the Library of the United States Department of Agriculture and later professor of the Rutgers Graduate School of Library Service, attended the "First Symposium on Machine Methods in Scientific Documentation," which I organized at Welch in March 1953. The chef of Johns Hopkins Hospital prepared a marvelous buffet luncheon. Later Ralph wrote and said, "Garfield, as a documentalist you make a great caterer!"

When Ralph was at the University of Hawaii, and I could not get to Hawaii to visit my son Stefan, who had been sent to the army hospital from Vietnam, Ralph went in my stead—a kindness I will never forget. And he instantly began bossing all the nurses around, telling them what they should be doing.

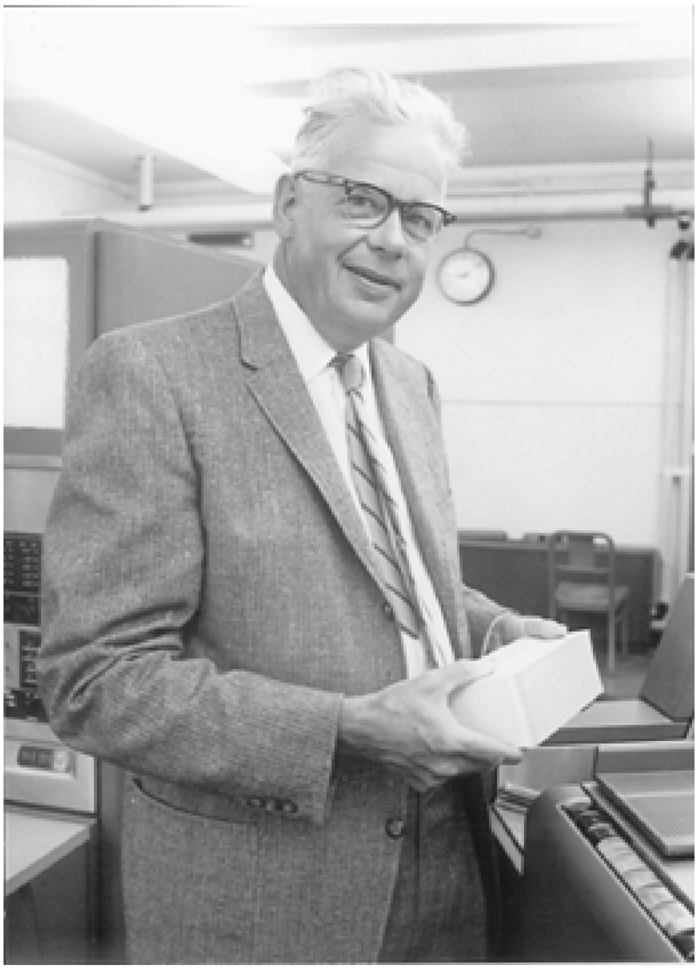
Another person I met at the Welch Project was John Mauchly, the co-inventor of ENIAC and UNIVAC. (The University of Pennsylvania recently celebrated the fiftieth anniversary of ENIAC.) When I came to Philadelphia, he and I became good friends. I had the sad task of doing a literature search for him about the blood disease to which he eventually succumbed.

Calvin Mooers, inventor of Zatocoding, was another pioneer I met at Welch (Garfield, 1997).

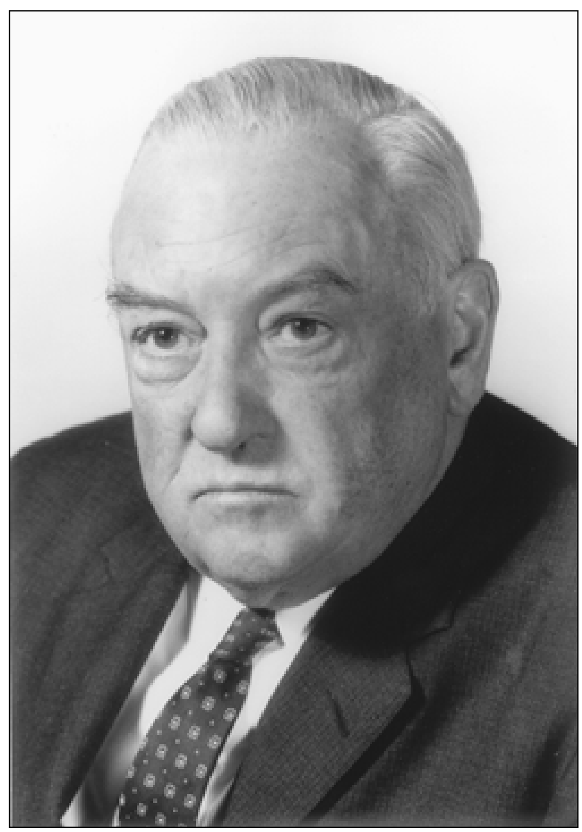
I never met the original chairman of the Welch Project's Advisory Committee of Honorary Consultants, Lewis H. Weed. He was succeeded by Chauncey D. Leake, who was dean of the medical school at the University of Texas at Galveston and the very first dean of an American medical school to hold a Ph.D. rather than an M.D. He was an authority on Leonardo da Vinci,



Joshua Lederberg.



James W. Perry.



Sanford V. Larkey.

Peter Luhn.

Verner Clapp. Courtesy Council on Library and Information Resources.

Ralph Shaw. Courtesy American Library Association Archives.

Calvin Mooers. Courtesy Helen Solorzano.

Chauncey D. Leake with his wife.

medical papyri, amphetamines, and California wines (Garfield, 1970, 1978a). He also served as president of the American Association for the Advancement of Science.

Williamina Himwich and Helen Field were employees of the Welch project when I arrived. Originally trained as a brain physiologist, Mina single-handedly indexed fifty years of the *American Journal of Physiology* after her retirement from the Welch Project. Helen Field was a medical librarian. Shortly afterward she married Judge Giles Rich of the U.S. Patent Court of Appeals, and they now live in Washington, D.C.

As project director, San Larkey reported directly to the head of the Army Medical Library, originally Colonel McNinch and then Frank Bradway Rogers. Brad Rogers is the person mainly responsible for developing the Medline system, described in my 1984 essay titled "Bringing the National Library of Medicine into the Computer Age: A Tribute to Frank Bradway Rogers" (Garfield, 1984). ISI is the continuing sponsor of the Medical Library Association Annual Award established in his name.

In that *Current Contents* essay I refer to Brad's colleagues at the Army Medical Library, which later became the Armed Forces Medical Library and then the National Library of Medicine in 1956: Seymour Taine*, editor of the *Current List of Medical Literature*; Estelle Brodman,* doyenne of medical historians and librarians (Brodman, 1954); Sam Lazerow, director of the library's acquisition division; Scott Adams,* medical bibliographer; Dave Kronick,* medical librarian (Kronick, 1976); and Robert Hayne, an assistant editor who also worked on the *Current List of Medical Literature*. He was a brilliant classics scholar. (In fact, I have found that the best indexers are classicists—not scientists.)

ISI Staffers

Bob Hayne later came to work at SmithKline & French in Philadelphia, where I was a consultant, and eventually to ISI as editorial director (Garfield, 1977). I am still in touch with his wife Virginia.

Another colleague I first met at SmithKline was Irving H. Sher. He later came to work for me as director of research. Not long after he came to ISI, Irv was joined by Art Elias from Wyeth. Art was the former editor of *JASIS*.

Bonnie Lawlor became a senior vice president of ISI, then left to go to UMI, and is now a consultant (Garfield, 1993a). She started as a chemical indexer and moved up to increasingly responsible positions at ISI. She remains very active in the American Chemical Society and ASIS.

Henry Small has been one of my closest colleagues for nearly thirty years. His seminal research on co-citation mapping is now a classical paper in information science (Small, 1973; Garfield, 1993b). His group at ISI has been responsible for hundreds of citation studies (Small, 1992). He was trained as a chemist and historian of science.

I met George Valdutz in Moscow about 1965 when he was at VINITI. Eventually he migrated to Vienna, and then to Rome, where I met him again and facilitated his emigration to the United States to work in ISI's chemical information research group.

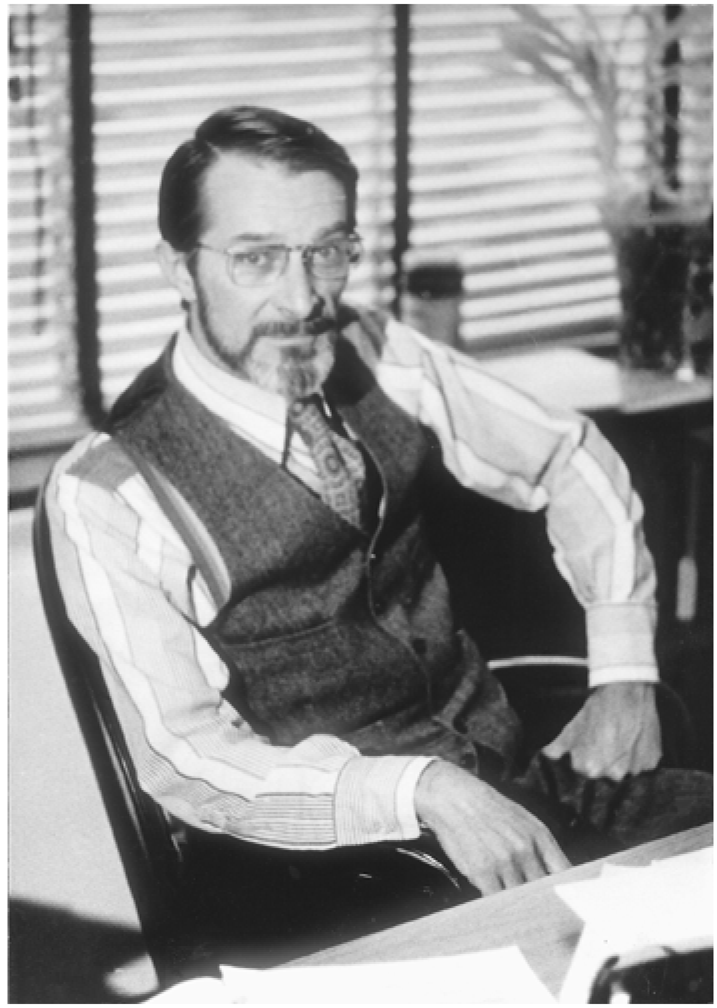
I met Sam Lazerow when he was at the Army Medical Library during the time of the Welch Project. He served in all three of the National Libraries (Agriculture, Medicine, and Library of Congress). He left the Library of Congress to become vice president of ISI (Garfield, 1972), but still maintained his Washington-Baltimore orientation by commuting from Baltimore every day. He became my closest personal friend and confidant. At ISI he was much beloved by the staff. In his memory ISI established the annual Lazerow lectures, which are held at a dozen library and information science schools.

Current Contents

In a short personal retrospective, space does not allow me to even mention all the people and events drawn from almost half a century of experiences. The origins of *Current Contents* seem to interest a lot of people. I could say that it all started in the Bronx, when I was about eight years old and happened to be living across the street from the Woodycrest branch of the New York Public Library. I was fascinated by titles. I used to sneak out of the children's section, where I was supposed to be, and literally read the titles of all the books in the rest of the library.

Many of you know that I started *Current Contents/Life Sciences* in the 1950s. But ASIS people rarely know that my very first contents page endeavor was in the field of documentation itself. I do not even have a copy of *Contents in Advance*, which I started at Welch in late 1952. Later on it got me into hot water with my boss, and he wanted it stopped. But it continued until after I left Columbia in 1954 and was taken over by Anne McCann. It ceased publication about a year later.

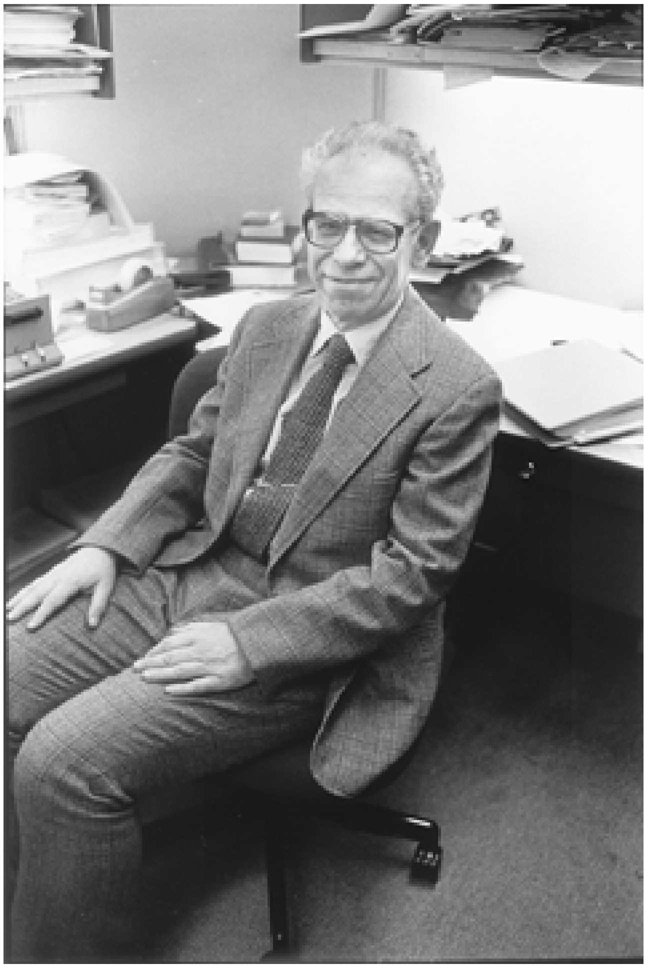
The Life Sciences edition of *Current Contents* became an official subscription service in 1958. Charlotte Studer was the special librarian at Miles Labs who gave me a contract in 1957. Contrary to general belief, the first *CC* title was in the management and social sciences, not the life sciences. The putative social sciences edition



Robert Hayne.

Bonnie Lawlor.

Frank Bradway Rogers.



George Vladutz.

Henry Small.

Sam Lazerow.

Eugene Garfield in front of the converted chicken coop in Thorofare, New Jersey, where ISI began.

Group including Irving Sher, Arthur Elias with Judy Leondar, Charles Bernier, and Bob Maizel.

Eugene Garfield and Pierre Vinken.

James Murray Luck. With permission, from the Annual Review of Biochemistry, *Volume 50 (1981), by Annual Reviews.*

started in 1954 with the name *Management's DocuMation Preview*. So I called my company DocuMation, Inc., which made Mort Taube very angry, since he had founded Documentation, Inc., in 1953. When we changed the name of *MDP* to *Current Contents* around 1955, I also changed the company name to Eugene Garfield Associates, Information Engineers. Mort was mollified, but the new name made the Pennsylvania professional licensing board very angry, because I did not have an engineering degree. So I dropped the subtitle "information engineers." Then in 1960 I changed the company name to the Institute for Scientific Information under the inspiration of the Russian VINITI. I bragged to a congressional committee that we could do with twenty people what they could do with twenty thousand (U.S. House Committee on Government Operations, 1974). (Later on I always enjoyed my visits to VINITI and respected their people a great deal.) Most of the twenty thousand were outside volunteer part-time abstractors who functioned much the way *Chemical Abstracts* did in those days. *CA* was their model.

I find that what people really want to know about the origins of *Current Contents* is how it was financed. I wrote an essay on this called "How It All Began: With a Loan from HFC" (Garfield, 1980). I have lost touch with Richard Gremling of Bell Labs, but he gave me a contract to produce a customized edition of *CC*. However, I did not have the sense to ask for an advance payment. In order to pay the printer, I had to raise $500 so I could deliver the first issue. None of the banks would lend me money, so I went to the Household Finance Corporation and got the money in about ten minutes. In those days, 6 percent per annum was normal bank interest. HFC charged 18 percent, but you did not have to keep the money for several months. So it turned out to be cheaper to pay 18 percent without service charges. And to get around the state law of a $500 maximum, I went to different branches of HFC. No networked computer records in those days!

What is significant about *Current Contents* vis-à-vis ASIS is somewhat bittersweet. *CC* has never been discussed seriously in the literature of information science. Even Brad Rogers said *Current Contents* was a "sop to appease the guilty conscience of doctors who don't read," or words to that effect. He had underestimated the importance of timeliness and simplicity, the essence of the original *Current List of Medical Literature*, which was started by Atherton Seidel during World War II. The problem with *Current Contents* is that it is so simple and utilitarian that it gives theoreticians very little to talk about. Current awareness is one thing—information retrieval is something else. It is telling that when I taught at the Moore School of Electrical Engineering at Penn in the 1960s, the engineers called *CC* an information retrieval tool!

Science Communication

My world of information science has been very broad in its scope. Indeed, I have not really touched on the field of science communication which has been my main concern—not just the indexing and abstracting of the literature, but also its reporting as exemplified by my newspaper, *The Scientist*. (I think it is reasonable to state that it was the first full-text journal available continuously on the Internet.)

The Scientist was the culmination of my thirty-odd years of writing over a thousand essays in *Current Contents*. That series appeared in *CC* because I realized early on that readers found going through *CC* a weekly necessary chore. The essays and cartoons provided a diversion—something concrete and amusing to read. It provided me the unique opportunity to deliver a weekly message to readers worldwide, but especially behind the Iron Curtain. I felt like a hero when I went to Czechoslovakia to receive an honorary degree. Hence my frequent allusion to the old adage that you are a prophet in a foreign land. The censors in Eastern Europe and China allowed what I had to say to go through, because *CC* was regarded as a bibliographic tool—not as a journal they had to censor.

Of the many science publishers I encountered in this vast world of science communication, I would like to mention Pierre Vinken, a neurosurgeon and editor—he may not be an information pioneer by ASIS standards. I met him in the 1950s when the Excerpta Medica Foundation was established. In time he converted this to a commercial enterprise, which has become one of the world's largest publishing conglomerates—Reed Elsevier.

There are dozens of other publisher friends I could mention, like Tom Karger, Gunter Heyden, Per Saugman,* and others. On the other hand, this reflection would not be complete without mentioning Robert Maxwell. We met in the late 1950s. Over the next thirty-five years he tried to acquire my company in one way or another, by hook or by crook. One day I will describe in detail about how he tried but failed. He was a diabolical, driven genius. Fortunately, the competitive world of publishing is full of other people who live and let live.

I have also not mentioned the dozens of science editors who are part of this scholarly publishing world, such as Stephen Lock, former editor of the *British Medical Journal;* Arnold Relman, former editor of the *New England Journal of Medicine*; Drummond Rennie, associate editor of the *Journal of the American Medical Association*; George Lundberg, former editor of *JAMA*; and Dan Koshland, former editor of *Science*, who recently received the Lasker Award. I first met Dan when I joined the board of *Annual Reviews* in 1979. *Annual Reviews* was founded in 1932 by J. Murray Luck, a biochemist who died several years ago at age ninety-three. Now *Annual Reviews* comes out in twenty-nine different editions. In 1979 *Annual Reviews* and ISI established the National Academy of Sciences Award for Scientific Reviewing (Garfield, 1979), and I used to write an annual essay about the winner.

Much of what I am telling you today has been covered in two oral history interviews—one by Arnold Thackray and the other more recently by Robert V. Williams, both for the Chemical Heritage Foundation ("Oral History Program," n.d.).

I am delighted that so many old timers have been able to come to this meeting, but there are many who are absent. Just last week I was able to contact Seymour Taine after a twenty-year hiatus. And I am glad to say that Estelle Brodman is still with us. The MLA (Medical Library Association) showed a remarkable video interview with her at the one-hundredth anniversary meeting in Philadelphia. I was delighted to hear that Fred Kilgour was able to come. I need not elaborate on his contributions to library and information science, in particular, OCLC (Online Computer Library Center).

As a concluding remark, let me say why I have always looked to ASIS as my home base. While I have been a member of many societies, including the ACS (American Chemical Society), ACM (Association for Computing Machinery), IEEE (Institute of Electrical and Electronics Engineers), ALA (American Library Association), SLA (Special Libraries Association), HSS (History of Science Society), NASW (National Association of Science Writers), and CBE (Council of Biology Editors), I have always considered ASIS to be my primary professional home base. ASIS provided a link between my friends and competitors in both the nonprofit and for-profit worlds—the link between academia, government, and industry. I hope that in the next several years as I serve as a board member and president that I can help build bridges in those areas so that ASIS can not only survive but grow and thrive in the new millenium.

Author's Note

Unless otherwise noted, photographs illustrating this talk come from my collection of photographs of information pioneers, which is available at http://garfield.library.upenn.edu/papers/heritagey1998.html. Because of time and space constraints, I cannot include all my photographs of pioneers. Asterisks in the text indicate that a photograph of the person is available on my Web page. There the reader will find images of still other worthies. I welcome the addition of even more.

References

Note that Eugene Garfield's Essays of an Information Scientist. *15 vols. (Philadelphia: ISI Press, 1977–1993) may be accessed in pdf form through the following site: http://garfield.library.upenn.edu/essays.html.*

Bernal, J. D. (1967). *Social function of science*. Cambridge, MA: MIT Press. (Original edition. [1939]. London: G. Routledge & Sons.)

Bonitz, M. (1997 December). The scientific talents of nations. *Libri, 47*(4), 206–213.

Brodman, E. (1954). *The development of medical bibliography.* Chicago: Medical Library Association.

Garfield, E. (1970). Calling attention to Chauncey D. Leake—renaissance scholar extraordinaire. *Current Contents,* April 22, 5–6. (Reprinted in *Essays of an Information Scientist*. [1977]. Volume 1, pp. 102–103. Philadelphia: ISI Press.)

Garfield, E. (1972). Introducing Samuel Lazerow, ISI's Vice President for Administration. *Current Contents*, November 1, 5–6. (Reprinted in *Essays of an Information Scientist*. [1980]. Volume 3, pp. 374–375. Philadelphia: ISI Press.)

Garfield, E. (1973). Introducing the Copywriter and ISI's subsidiary, Selective Information Devices, Inc. (SID). *Current Contents,* May 2, 5–8. (Reprinted in *Essays of an Information Scientist*. [1977]. Volume 1, pp. 438–442. Philadelphia: ISI Press.)

Garfield, E. (1975). The "obliteration phenomenon" in science—and the advantage of being obliterated! *Current Contents,* December 22. (Reprinted in *Essays of an Information Scientist*. [1977]. Volume 2, pp. 396–398. Philadelphia: ISI Press.)

Garfield, E. (1977). To remember my brother, Robert L. Hayne." *Current Contents,* August 22, 5–6. (Reprinted in *Essays of an Information Scientist*. [1980]. Volume 3, pp. 213–214. Philadelphia: ISI Press.)

Garfield, E. (1978a). To remember Chauncey D. Leake. *Current Contents,* February 13, 5–15. (Reprinted in *Essays of an Information Scientist*. [1980]. Volume 3, pp. 411–421. Philadelphia: ISI Press.)

Garfield, E. (1978b). To remember Ralph Shaw. *Current Contents,* June 5, 5–11. (Reprinted in *Essays of an Information Scientist*. [1980]. Volume 3, pp. 504–510. Philadelphia: ISI Press.

Garfield, E. (1979). The NAS James Murray Luck Award for Excellence in Scientific Reviewing: G. Alan Robison receives the first award for his work on cyclic AMP. *Current Contents,* April 30, 5–9. (Reprinted in *Essays of an Information Scientist.* [1981]. Volume 4, pp. 127–131. Philadelphia: ISI Press.)

Garfield, E. (1980). How it all began: With a loan from HFC. *Current Contents*, January 21, 5–8. (Reprinted in *Essays of an Information Scientist*. [1981]. Volume 4, pp. 359–362. Philadelphia: ISI Press.)

Garfield, E. (1982a). J. D. Bernal—the sage of Cambridge. 4S Award memorializes his contributions to the social studies of science. *Current Contents,* May 10, 5–17. (Reprinted in *Essays of an Information Scientist*. [1983]. Volume 5, pp. 511–523. Philadelphia: ISI Press.)

Garfield, E. (1982b). More on the ethics of scientific publication: Abuses of authorship attribution and citation amnesia undermine the reward system of science." *Current Contents,* July 26, 5–10. (Reprinted in

Essays of an Information Scientist. [1983]. Volume 5, pp. 621–626. Philadelphia: ISI Press.)

Garfield, E. (1984). Bringing the National Library of Medicine into the computer age: A tribute to Frank Bradway Rogers. *Current Contents,* January 9, 5–11. (Reprinted in *Essays of an Information Scientist*. [1985]. Volume 7, pp. 5–11. Philadelphia: ISI Press.)

Garfield, E. (1993a). Bonnie Lawlor's silver anniversary at ISI—A tribute. *Current Contents,* April 12, 3–5. (Reprinted in *Essays of an Information Scientist*. [1993]. Volume 15, pp. 280–282. Philadelphia:ISI Press.)

Garfield, E. (1993b). Co-citation analysis of the scientific literature: Henry Small on mapping the collective mind of science. *Current Contents,* May 10, 3–13. (Reprinted in *Essays of an Information Scientist*. [1993]. Volume 15, pp. 293–303. Philadelphia: ISI Press.)

Garfield, E. (1997, March 17). A tribute to Calvin N. Mooers, a pioneer of information retrieval. *The Scientist, 11*(6), 9.

Goldsmith, M. (1980). *Sage: A life of J. D. Bernal*. London: Hutchinson.

Goldsmith, M. (1995). *Joseph Needham: 20th-century renaissance man*. Paris: UNESCO Publishing.

Kronick, D. A. (1976). *History of scientific and technical periodicals*. Second edition. Metuchen, NJ: Scarecrow Press.

Lederberg, J. (1978). Digital communications and the conduct of science: The new literacy." *Proceedings of the IEEE, 66*(11), 1314–1319.

Merton, R. K. (1965). *On the Shoulders of Giants*. Chicago: University of Chicago Press. (Original edition. [1965]. New York: Free Press).

Merton, R. K. (1968a). The Matthew effect in science. *Science, 159*(3810), 56–63.

Merton, R. K. (1968b). *Social theory and social structure*. New York: Free Press.

Oral History Program at http://www.chemheritage.org.

Proceedings of the International Conference on Scientific Information, Washington, DC, November 16–21, 1958. (1959). 2 volumes. Washington, DC: National Academy of Sciences, National Research Council.

Small, H. G. (1973). Co-citation in the scientific literature; a new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24,* 265–269. (Reprinted in *Essays of an Information Scientist*. Volume 2, pp. 26–31. Philadelphia: ISI Press.)

Small, H. (1992). Citation classic: Cogitations on co-citations. *Current Contents*, March 9, 20.

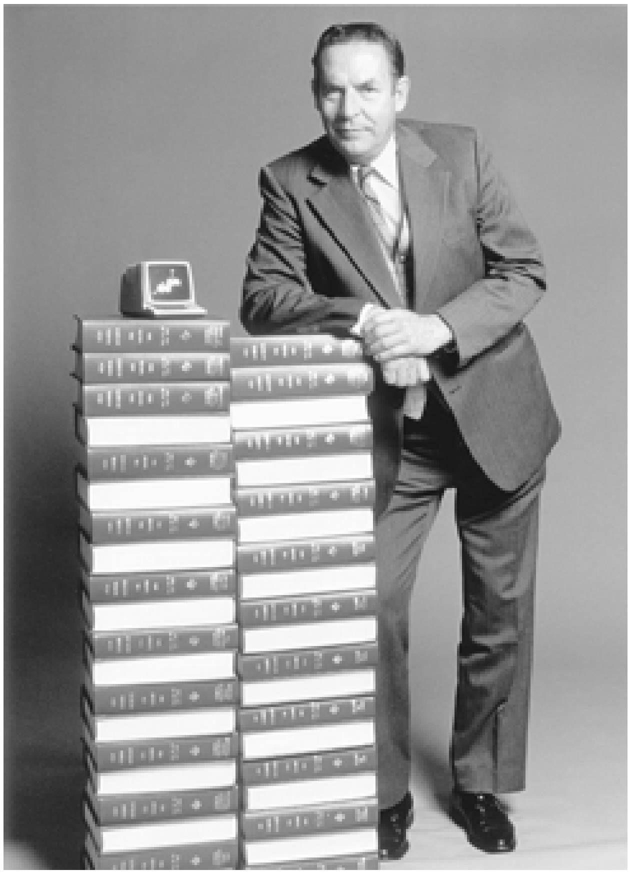
U.S. House Committee on Government Operations. (1974). *Federal information systems and plans—Implications and issues.* Hearings before the Subcommittee of the Committee on Government Operations, 93rd Congress, second session. Part 3, 883–901. Washington, DC: Government Printing Office.

Zuckerman, H. (1996). *The scientific elite: Nobel laureates in the United States*. New Brunswick/London: Transaction Press.

Pioneers' Reminiscences

When the conference planning committee began its work, one of its principal objectives was to involve the pioneers of information science in as many aspects of the conference as possible. The committee invited about thirty pioneers to the conference and a gala dinner in their honor, for which each was asked to prepare brief remarks on a "memorable moment" in his or her career (or simply bring greetings). Twenty pioneers were able to attend the conference and make such presentations. Their remarks were so interesting and varied that a request was made both to those attending and to those not able to come to include their stories and photographs in a special "scrapbook" located on the Pioneers of Information Science Web site (www.asis.org) and in this volume. Unless otherwise noted, photographs illustrating these reminiscences were supplied at our request by the pioneers. (For a short period, Robert V. Williams at bobwill@sc.edu will continue to post late responses on the Web page.)

The following pages reflect the individuality of the information science pioneers. They contain a few historical gems, speculations about the future of the field, and insights into what motivated the careers of the pioneers. Those individuals who responded to our specific request for a "memorable moment" provided stories that are not only fascinating in detail but also revealing of the larger scene that was developing as the field of information science.



Reprinted with permission of Business First, *Business First of Columbus, Inc.*

Dale B. Baker

World's First Electronically (Computer) Published Scientific Journal

"Both science and history are theories drawn from actual events; presenting both successfully is difficult."

Mary Jo Nye, Berkeley

Research and Development at Chemical Abstracts Service (CAS) took off rapidly in 1959 when G. Malcolm Dyson came from England to become its director. In April 1959 Dyson invited Hans Peter Luhn, director of IBM Systems Development Division, to visit CAS. Dyson wanted to talk with Luhn about the software he had recently developed, keyword-in-context (KWIC), and an automated indexing technique that he had reported on at an ACS Chemical Literature Division meeting in Atlantic City. Over lunch with Luhn, Dyson and I decided to test the KWIC technique in *Chemical Abstracts.* Our users had been clamoring for a fast abstracting and indexing service "of the top 10 percent of the world's most important chemical research literature."

But CAS had no computer! We immediately leased an IBM 1401, but as the demand for computers was great, it took twenty-one months for IBM to deliver.

Until 1959 the ACS board of directors had an unwritten policy not to seek or accept government grants or contracts (for fear of government influence or control). Thus the ACS board had to approve our request for the needed funding to support R&D projects from the newly established Office of Scientific Information of the National Science Foundation. We submitted a proposal to NSF for $112,000 to establish a pilot plant operation and publish four test copies of this new concept periodical, *Chemical Titles.* The request included money for market research. Thankfully, the NSF grant was promptly approved.

Because our IBM 1401 had yet to arrive, we had to hand carry punched cards to Poughkeepsie, New York, to run on the IBM software and computers to compose the initial issues of *Chemical Titles.* Two test issues of the "quick and dirty" *Chemical Titles* (*CT*) were electronically produced in 1959, and a third issue was produced in 1960. All were favorably received, so the fourth test copy was judged not to be needed, and the project came in under budget.

As part of the market research, potential subscribers were asked if they would subscribe at $25, $30, or $35 per annum. Of course, the majority opted for a $25 subscription price. However, the ACS board did not accept staff's recommendation of $25, and the price was fixed at $30. I was quite dejected; this was my first experience of not getting board acceptance of staff-developed recommendations. While it may seem to be just a $5 difference today, all chemists know that 20 percent is considered significant in any analysis. Fortunately, *CT* started regular weekly publication in January 1961 and was a success. The publication continues today. But the base subscription price is $610 per year now or $240 for ACS members.

Also in 1961 information scientists at Purdue University and Olin Corporation requested *CT* tapes to experiment on off-line searching for the subject profiles of their research scientists. These experiments were to run for eighteen months, and the results were to be reported back to CAS. By the end of 1964, there were some 116 leases of *CT* tapes being sent out on a regular basis. The information experts using the *CT* tapes met at CAS semiannually for two days starting in 1964 in workshops and seminar-type meetings. These same specialists founded the Association of Scientific Information Dissemination Centers (ASIDIC) in 1968.

Thus, was born the world's first periodical to be organized, indexed, and composed by computer. It was also the beginning of the computerization of all operations and information services at CAS.

Everett Brenner

The Day I Faced Technology

Brenner's Law

Determine the best system you can foresee before designing the system you can afford.

As long ago as the early 1960s, as director of the American Petroleum Institute's Central Abstracting and Indexing Service, I had to consider technology and the coming of the Information Age. Computer specialists from various petroleum companies met with me to discuss how technology might help in automating our abstracting and indexing endeavors, in aiding users search more efficiently, and in reducing costs. I went for broke, envisioned the ideal, and asked whether computers could achieve what may even have seemed outlandish to many who were present. In light of what computers can do today, it's possible my vision may have been conservative. In any case, the computer specialists said they could accomplish all I asked for at the time, but much of what I asked for would be very expensive.

From that moment on I have never had difficulty living with and using whatever technology offered. I realized then and there that I could govern technology; technology would never govern my final decisions. I was able to envision the ideal, but knew I would have to be able to compromise for what was practical, what I could afford and what would be competitive. This is not an ideal world and one may never be reached. I doubt whether total intelligence for information retrieval can ever be achieved, but it is important to strive for it. For me, it's been a top-to-bottom-to-top approach. Go for broke, know what you want to achieve, compromise when necessary, and then start all over again as more and more technology becomes available. The bottom-to-top approach keeps one focused only on the next step—not the highest and the best. With that approach the American Petroleum Institute's service was to be one the first to go online, instituted one of the earliest text editing systems, and developed a machine indexing system that retrieves 70 percent of the required controlled terms automatically.

Helen L. Brownson

Organizing Scientific Information after *Sputnik*

As I think back over the early days of information science, when I was at the National Science Foundation, the memorable "moment" that first comes to mind occurred right after the 1957 launching of *Sputnik*. I was startled one evening to hear Eric Sevareid in his 11 P.M. radio broadcast describe the Soviet Union's All-Union Institute of Scientific and Technical Information and its coverage of the world's scientific and technical literature. He contrasted it with the Office of Technical Services, in the Department of Commerce, and its abstract journal, *U.S. Government Research Reports*. He apparently did not know that in the United States, scientists and engineers relied on abstracting and indexing services covering their fields, such as *Chemical Abstracts*, *Biological Abstracts*, and *Index Medicus*.

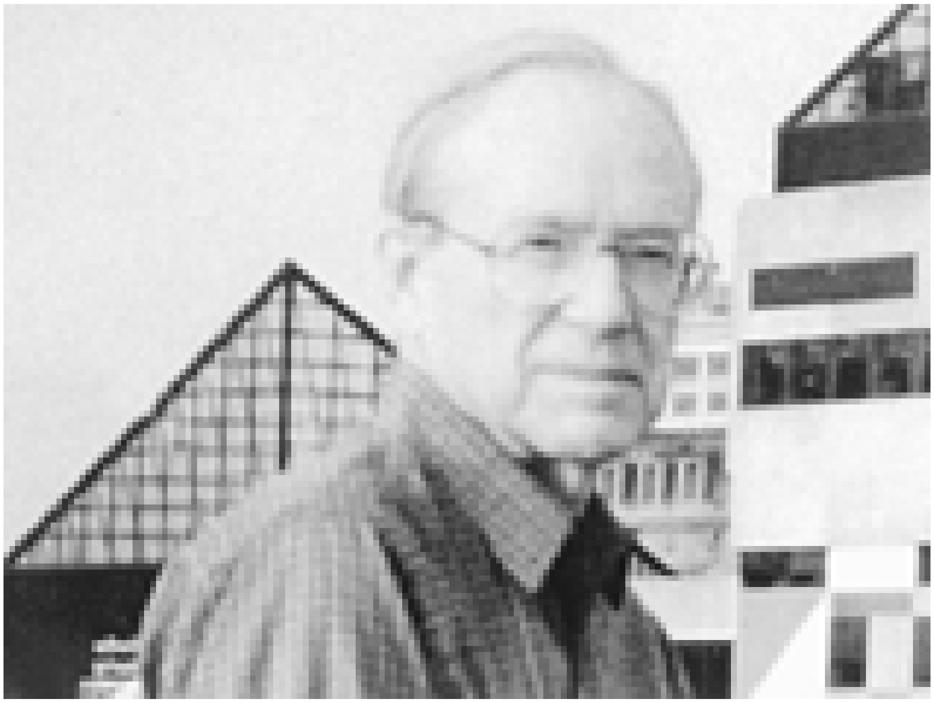
It seemed to me that this broadcast, and other reactions to the demonstrated technical achievement of the Soviets, contributed to a change of climate surrounding information science and also the NSF's efforts to develop programs to facilitate access to scientific literature. Thus 1958 became the busiest year in my recollection.

The president and his Science Advisory Committee, headed by James R. Killian, took an interest in the science information problem. The committee formed a special subcommittee, headed by W. O. Baker, vice president of Research at Bell Telephone Laboratories, to look at the problem of improving access to scientific literature. This subcommittee turned to the NSF staff for certain kinds of help and information, and we spent much time that year preparing special reports on the current problems and possible solutions.

Then, with the aid of the subcommittee, the Science Advisory Committee issued "Improving the Availability of Scientific and Technical Information in the United States" on 7 December 1958. At the same time the White House issued a press release stating that "the president today approved a plan designed to help meet the critical needs of the nation's scientists and engineers for better access to the rapidly mounting volume of scientific publication." The president directed that NSF take the lead in bringing about effective coordination of scientific information activities within the federal government. Our mission was thus made clear.

NSF announced the establishment of a Science Information Service on 11 December 1958, its objective was to extend the foundation's existing science information programs in order to carry out the president's objective. To oversee development of this expanded effort, a Science Information Council was named with representation from federal agencies and private scientific organizations.

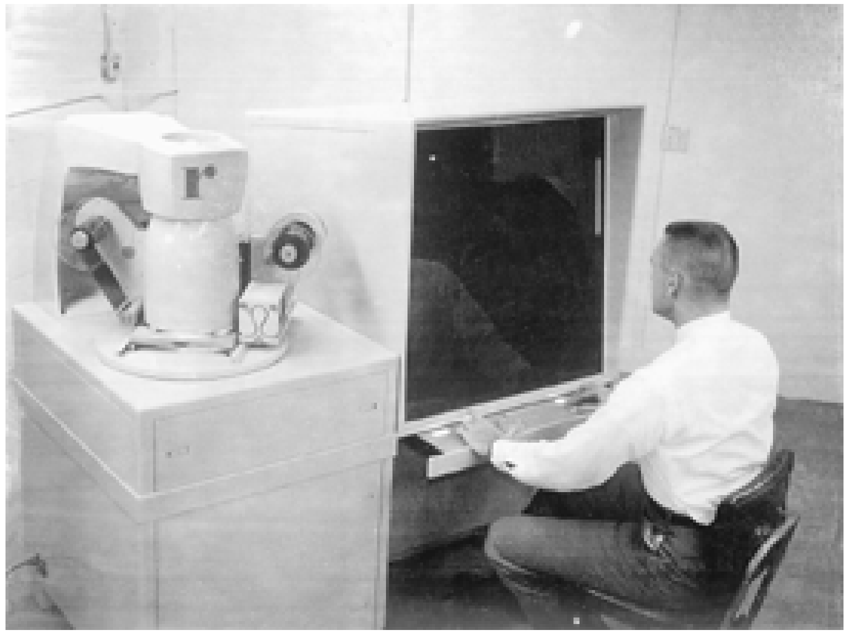
I might add that this memorable year was made still busier by NSF's participation in organizing the large International Conference on Scientific and Technical Information, which was held in Washington, 16–21 November 1958. It was sponsored jointly by NSF, the National Academy of Sciences–National Research Council, and the American Documentation Institute (predecessor of the ASIS). The conference was conceived three years earlier by members of ADI. I served on the program committee with responsibility for organizing and summarizing Area 1 (Literature and Reference Needs of Scientists). The conference addressed seven areas altogether. The proceedings were published in two volumes by NAS-NRC in 1959.



Lee Burchinal

I consider the publication of the first issue of *Research in Education*, the monthly announcement bulletin of the Educational Resources Information Center (ERIC) system (now *Resources in Education*), that contained output from the ERIC clearinghouses the most satisfying and significant event in my career. The process leading to this event began in the spring of 1965, when with a few staff and a consultant, Fred Goodman of the University of Michigan, I decided to stake my career on a novel and risky design for the ERIC system. I agreed to vest responsibility for document acquisition and processing in the hands of then inexperienced staff at subject-oriented clearinghouses, primarily at universities throughout the country. The decentralized design was contrary to the conventional tightly controlled centralization of document acquisition and processing under one roof. But given the decentralized American educational system, I felt we had to adopt a comparable decentralized design.

To implement this process, we had to arrange for two other crucial elements in a decentralized network: the ERIC facility to receive output from the clearinghouses and produce a computer tape for printing *Research in Education* (*RIE*) at the GPO and the ERIC Document Reproduction Service (EDRS) to reproduce in microfiche and hard copy the documents announced in *RIE*. The first EDRS contract was awarded in November 1965, and the first ERIC facility contract in May 1966. Both were with for-profit firms—another break from the practice of in-house government production. In the spring and summer of 1966 contracts were awarded to establish eighteen subject-based clearinghouses. In July 1967 the first issue of *RIE* with document resumes from the clearinghouses appeared, marking the end of the beginning of the ERIC system. Other ERIC successes followed, including the production of *Current Index* to *Journals in Education*, to cover the journal literature; becoming one of the first federal systems available online; and seeing usage surge. But the moment in July 1967 when we saw the first tangible monthly output of the ERIC system remains as my fondest memory of my information career at the Office of Education and later at the National Science Foundation.



Robert Chartrand at the first photographic interpretation console.

Robert Lee Chartrand

Mapping a Virtual Country

In a career spanning more than forty years, there have been many special moments to recall and savor. Perhaps paramount was the opportunity to play a significant role in the development of "Subsystem I" (for Intelligence) of Air Force project 117L. The purpose of this endeavor was to create a pilot Data Systems Laboratory (DSL) in Littleton, Colorado, that would serve as the testing site for processing the "take" from the first, and as yet un-launched, spy-in-the-sky satellite.

As a Member of the Technical Staff at Thompson-Ramo-Wooldridge (T-R-W), the subcontractor of Lockheed, it was felt that my contribution could be threefold: I had experience as a multisensor analyst (including postings at the U.S. Navy Photo Interpretation Center (NPIC) and OP922Y1 at the National Security Agency); I understood Soviet bloc demography and politics; and I was familiar with a range of mapping and photographic resources.

Because existing coverage of Communist countries was highly classified, it was believed that simulating these geographic areas could provide useful lessons in the future processing of photo and electronic intelligence (ELINT) data. Therefore, a country called Slavia was created as a fictional counterpart to the Soviet Union; it featured all aspects of a real country—geography, armed forces, economy, and so forth.

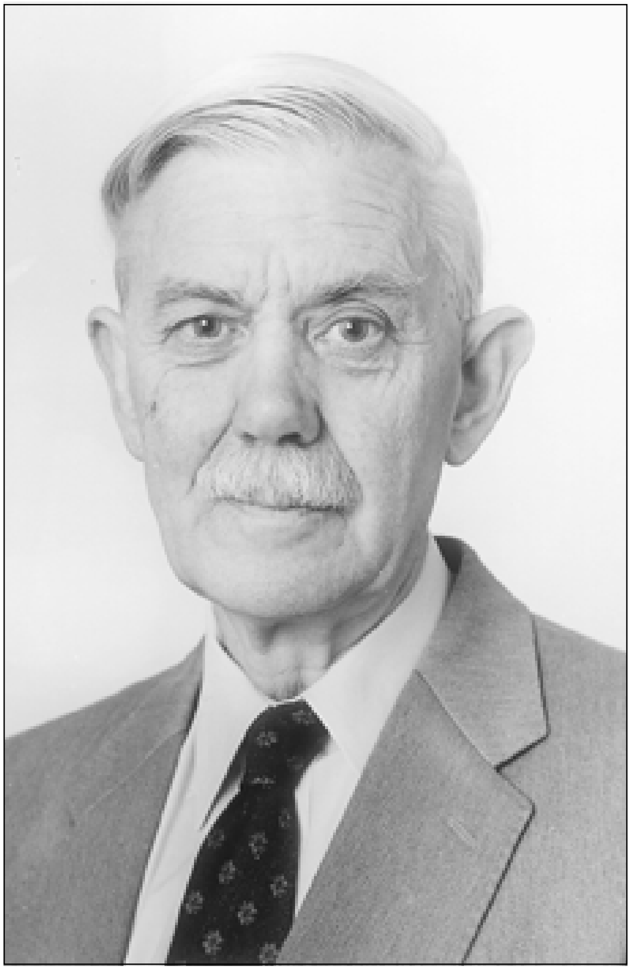
The hardware support system that was built by a group of vendors (including Itek and Houston-Fearless) revealed a willingness on the part of T-R-W to break new ground. It included photo interpretation and ELINT consoles, a twin-screen display analysis console, a central store to house and manipulate 70×100-mm film chips, a group display unit, a family of photographic and format conversion devices, and a new "polymorphic" computer (the RW-400).

In using Slavia as a realistic, if fictional, database, certain key components were developed:

- A basic description of this country, including the details of military resources, geographic components, economy, and its people.
- A satellite camera for taking pictures of large photomosaic boards representing "ground truth" in Slavia.
- Generation of these mosaic boards, of requisite quality and scale, featuring target inserts—airfields, missile sites, manufacturing plants—which could be updated.
- In-series Slavian maps, similar to real products from U.S. mapping agencies.
- Various useful scenarios to exercise this elaborate database.

Working closely with assigned Air Force intelligence specialists who served as the test bed cadre, our project team—computer and microform systems' specialists, language experts, human scientists, and photographic personnel—monitored and evaluated the DSL Electronic Center operations and developed demonstrations of these sophisticated devices and systems.

In urging and approving my acceptance of this opportunity, my leader and friend at NPIC, Arthur Lundahl, had viewed this as an "absolutely unique and crucial" project. In retrospect, it simply reinforced our shared belief that "where there is no vision, the people perish."



Eric J. Coates

Downsizing the Hunch Element in Subject Indication: My First Meeting with Ranganathan

Looking back to ask myself what was the most significant episode in my professional life, I instantly recall my first meeting with S. R. Ranganathan. This meeting completely turned around my thinking about classification in library and information science and set the direction of my subsequent career. The year was 1950. I was one of the team appointed to launch the British National Bibliography (BNB). If I had any particular expertise that led to my appointment, it was not in classification but, rather, in what is now called descriptive cataloging, where I had some experience in tutorial work. However, I had read the classification literature recommended for students and formulated the view that library classification was largely a hotchpotch of folklore-like precepts without any adequate connective principles behind them. To my initial consternation, after a few months at BNB I was asked to head up the subject indication side of the project. Overseeing my colleagues' classification decisions soon confirmed my worst fears: I could see no overarching principles to underpin consistency either between the decisions of individual indexers or between a variety of different solutions to parallel problems. Such agreed rules as there were proved woefully inadequate to resolve many daily recurring problems.

Then came the occasion when Ranganathan visited the BNB office. He agreed to be questioned and to give advice on a list of the practical dilemmas in all their often intricate detail that we had met in the first months of the BNB operation. He dealt with our challenging queries quietly in a direct head-on manner and with an economy of words in which neither sidestepping nor obfuscation could have any place. He spoke slowly with gaps between clauses and sentences, making time for us to absorb fully his points. When he was unable to give a ready response to a query, he would suggest that we solve the problem together. It was a truly enlightening experience to accompany him step-by-step in his extempore thinking. In retrospect I feel it is not possible to exaggerate the impact that this first meeting with Ranganathan had upon me. It moved me from cautious skepticism to confidence that the search for a coherent and communicable rationale for the practice of subject indication was no chase after a will-o'-the-wisp. Like all mortals, Ranganathan had a few blind spots, but his facet analysis was a gigantic step in the search. Some decades later this first meeting became the inspiration for my work in connection with the second edition of the *Bliss Bibliographic Classification* and the *Broad System of Ordering*. These are demonstrations of what coherent general classifications can offer to the age of mechanized retrieval and of the Internet.

Pauline Atherton Cochrane and Hubert H. Humphrey chat at an ASIS meeting. Courtesy American Society for Information Science.

Pauline Atherton Cochrane

A Report to the Institute of Physics

The 1960s and 1970s were a special time for Science Information Systems. The leaders at the National Science Foundation (Burton Adkinson and Helen Brownson to mention only a few) provided more than financial support. They provided opportunities for researchers and practitioners to meet and discuss how they could learn from each other, work together, and advance the state of the art. The terms *mission-oriented* and *discipline-oriented information systems* meant something then, and the idea of cooperation, collaboration, and networking were seen as necessary and desirable if we were to improve information systems for scientists in this country and elsewhere.

Because the two major discipline-oriented abstracting and indexing services for the English language community had divided the pie in the early 1900s, it was necessary for me to work very closely with the Institute of Physics in London as well as the American Chemical Society in Columbus, Ohio. There was no *Physics Abstracts* published in the United States. What I heard from the physics community could be summarized as "*Chemical Abstracts* covers that part of physics that a chemist can understand and *Nuclear Science Abstracts* covers that 'small' field called nuclear physics. For the rest of physics we have to go to the British publication and it needs to cover physics better than it does."

My task then, from 1961 to 1966, while assistant director of the Documentation Research Project at the American Institute of Physics, was to ensure improvements in the coverage of physics research by the world's major abstracting and indexing services and, if possible, to draft the requirements and begin the development of a new, computer-based reference retrieval system that physicists could trust and would use.

The memorable moment I want to share happened when I submitted a confidential report to the Institute of Physics on the coverage of *Physics Abstracts* for 1961. (The public report appears as the publication AIP/DRP PA1 (1964), *The Journal Literature of Physics; A Comprehensive Study Based on Physics Abstracts*, by Stella Keenan and Pauline Atherton.)

I used unit-record equipment to perform a bibliometric study of the 20,287 abstracts in *Physics Abstracts* (*Science Abstracts*, Section A). This meant sorting 20,287 cards over and over to analyze several characteristics of these information items—where published, language, number of authors, subject placement in *Physics Abstracts* classification outline. (Today you could easily determine these data from online searches.) Up to that time no one could answer such questions as how many articles appear in *Physics Abstracts* under the category Solid State Physics, from how many journals, from how many countries, and in how many languages? Does *Physics Abstracts* cover the publications of the American Institute of Physics cover to cover? What is the time lag after publication before the abstracts for these articles appear in *Physics Abstracts*? Does *Physics Abstracts* cover all the books on physics which have been reviewed in AIP journals? If not, why not?

I sent my confidential report to London a month before I traveled there in August 1962 to work out ways we could use the report's findings to improve *Physics Abstracts*. At the first meeting I could feel the veiled antagonism and coolness from the assembled staff, including Bernard Crowther, the editor of *Physics Abstracts* and even the director of the Institute of Physics. Each staff member had a large report in hand, but it was not my report! It was their response to my report, attempting to renounce findings on book coverage and on time lag and journal coverage details (e.g., 4,511 of the 1961 issues covered pre-1961 publications; 8 percent, pre-1960!). I was all alone in this room of experienced abstractors and indexers and respected physicists. I had no credentials to offer that would upgrade my report—after all I was just a librarian with no scientific background.

But one thing I did have that they didn't—data. Data that their data couldn't refute. They really did not have, at that time, an easy way to determine time lag or coverage of journals and books that did not come to their offices directly from publishers. They occasionally went over to the British Museum Library if they noticed that their copy of an important new journal had not arrived in the mail. So after the initial skirmish we agreed to try and improve their acquisition procedures, to investigate scanning the input at the Boston Spa Science Library to improve coverage of foreign-language physics journals and to improve indexing procedures with automated techniques. Remember that this was before INSPEC. Maybe in some small way I helped get INSPEC into high gear by encouraging the Brits to do what we in the United States would not do any time soon. Our approach at the time was not to build a new system for physicists unless the existing ones couldn't be improved.

I also worked with *Nuclear Science Abstracts* and *Chemical Abstracts* in the United States. We all felt part of a team that had government support, professional society interest, and an obligation to do more for their members, and the possibility of implementing any research and development efforts that appeared likely to improve the current situation. My six years at AIP, working with journal editors, information center personnel, librarians, and other researchers was a most exciting time, in this country as well as in Europe, but that initial meeting at *Physics Abstract* is one I will always remember as particularly breathtaking.

Courtesy American Society for Information Science.

Melvin S. Day

A Moment in Time

For generations, scientific journals and technical books have been principal media for communicating scientific, engineering, and medical information. World War II gave the technical report a life of its own—even within an environment of controlled access to security-classified information. Nowhere was this limitation more pronounced than on the nation's Manhattan Project (atomic bomb project) from 1942 to 1946. Compartmentalization of information was a way of life for all of us on the project. I knew the details of what I was doing, and I knew what my staff was doing. I did not know, nor was I supposed to know, what my immediate management was doing or what my colleagues in other laboratories were doing. The only reports I wrote were to my supervisor. The only reports I read were from my own staff.

In 1946 Congress voted to establish the U.S. Atomic Energy Commission (USAEC) effective 1 January 1947 and to transfer to that civilian agency all existing duties of the Manhattan Project being operated by the Army Corps of Engineers. During the war everything on the project was classified. In 1946 the army decided to declassify as much of this information as possible for the education of the public about atomic energy as well as for the use of the soon-to-be-established USAEC. The Army Corps of Engineers had not issued policies covering the preparation of its Manhattan Project technical reports. As a result, when these technical documents were declassified many did not carry any of the information (e.g., author's name, dates, pagination) that we take for granted today. Concurrent with undertaking the major declassification effort, the army established a small technical information documentation program to organize the information materials that were declassified and to start cataloging, abstracting, indexing, publishing, announcing, and making them available to the public. Alberto F. Thompson headed the technical information program, and he asked me to join him before my discharge from the army. Bernard Fry became the chief librarian and he selected Israel A. Warheit to direct the program for cataloging, abstracting, and indexing the information materials that would be turned over to the USAEC for its stewardship.

When the USAEC was formally established, a large number of documents were being declassified. Following a time-honored library practice, a set of catalog cards for each report distributed by the central technical information office in Oak Ridge, Tennessee, was distributed to all Atomic Energy Commission National Laboratories and Atomic Energy Commission contractors. Each technical document was indexed in-depth with as many as sixteen subject cards included in a single set. Within a relatively short time large backlogs of unfiled cards began to accumulate across the country. The card recipients were literally drowning in catalog cards, and the USAEC discontinued its practice of issuing catalog cards covering reports. In place of the cards the USAEC issued a monthly current awareness tool with indexes, *Abstracts of Declassified Documents*. In the late 1940s this journal was expanded and became *Nuclear Science Abstracts* with world coverage of all unclassified nuclear science reports, published articles, books, and handbooks.

Each monthly issue had four separate indexes: subject, author, corporate author, and report number. Quarterly, semi-annual, and annual indexes were issued shortly after each calendar period that they covered.

An important moment in time for me was the development of the atomic energy technical report from its elementary form covering government war-time programs in the 1940s to a highly used and valuable medium for communicating scientific and technical information. At the same time (fifty years ago) the USAEC developed its precedent-setting abstract journal, *Nuclear Science Abstracts*, as both a current awareness announcement tool and an in-depth finding tool. In today's electronic world, the production systems that we developed, although effective at the time, would be out-of-date and old-fashioned. The story of how the USAEC accomplished this in the 1940s with the use of only electric typewriters and card sorters is a fascinating moment in time but must be left for another day.

Courtesy Chemical Heritage Foundation.

Thomas F. Deahl

The Origin of Electronic Binary Computation

This page is much too short to recount all the exciting moments during a career in the field of information science. In my case it was less empirical science than applied systems. I would not deny that we information system engineers had our experimental moments. Indeed, I proudly worked shoulder to shoulder with the folks who brought you the Y2K problem. One anecdote that may be of interest is the origin of the idea for electronic binary computation, the fundamental building block of today's computers.

I had the good fortune to know John Mauchly who, with J. Presper Eckert, invented the general-purpose electronic digital computer. Mauchly had a long-standing interest in weather forecasting. He was certain that long-range prediction of weather conditions could be done if he could calculate the interaction of enough variables fast enough. In the 1930s, while a professor of physics at Ursinus College, he wondered if some sort of electronic calculator could be devised that would give him the speed he needed. One Saturday, while shopping for screw-type house electrical fuses, he encountered a new fuse product by either Westinghouse or General Electric—I can't remember which. At any rate, when an electrical circuit had blown, a tiny neon lamp across the top of the fuse would light. The purpose was to help you find the blown fuse. Mauchly had an epiphany. He observed that this new fuse signaled whatever state the circuit was in. Current was either flowing or not. The neon lamp was either on or off. Zero or one. He bought a gross of these fuses and took them back to his laboratory where he built a circuit that could calculate sums by screwing and unscrewing fuses. The rest, as they say, is history.

Robert A. Fairthorne

A Congress at Harvard

My transition from the lower levels of aeronautical research to information science took some decades—of chance rather than of choice—in the Royal Aircraft Establishment at Farnborough, England. During these years organizing the series of tools for the previous tasks became the main task. There came a time when either computation or information was the main task. The changeover to information was consolidated in 1950, after a visit to United States for the International Congress of Mathematicians at Harvard.

My brief was to report on the congress, to visit some pioneering computer projects, and to find out about an interesting type of superimposed coding devised by Calvin Mooers. The Congress Proceedings also contained a paper by Mooers, "Information Retrieval Viewed as Temporal Signaling," which, apart from being of seminal importance, was the first appearance between hardcovers of Calvin's coinage *information retrieval*.

Personally this visit was memorable because it was the first time I had flown as a passenger rather than as an observer. Even more so because it was my first visit to the United States from a Britain still afflicted with strict rationing and acute shortages. Above all, it included my first visit to a U.S. home, as opposed to a hotel—the home of Calvin and Charlotte Mooers. Thus began a friendship that lasted till Calvin's death a few years ago.

Though this was a milestone in my progress toward information science, the journey was not complete nor a new one begun, till I resigned from the establishment in 1963.

Douglas Foskett

Librarianship to Information Science

After six years of war duty, I rejoined the Ieford Public Library service in 1946 and set about completing my F.L.A. This service had a good tradition of assistance to readers, and when I joined the Metal Box Company in 1948, I soon realized how the skills required for a scientific and industrial research information officer depended on the basic techniques of librarianship, notably classification and cataloging. The enhancement of these led to the development of higher levels, in literature searching, and, more particularly, in current awareness service and selective dissemination of information.

The Research Division Library of Metal Box Co. served all the factories as well as scientific staff, and we published an *Information Letter,* which included readable abstracts of current sci-tech literature in a form appropriate for busy executives and factory superintendents. Presenting scientific and technical information and building on librarianship skills gave me the magic opportunity to be among the pioneers of the emerging paradigm that came to be known as "information science," and I meant to proclaim this in the title of my book *Information Service in Libraries 1958*.

Meeting with S. R. Ranganathan in 1948 gave me a new view of classification as facet analysis plus traditional generic analysis, and I applied this in schemes for packaging, occupational safety and health, and education. This experience has suggested to me that facet analysis applied to any subject can reveal hitherto uncoordinated concepts—for example, materials and processes—and thus offer an indication of possible areas of future research. This could be a unique information science to the World Wide Web.

I cannot recall any special moment when I exclaimed "Eureka!" and rushed, like Archimedes, naked into the street. The opportunity to be present when new ideas were around and old methods were being adapted and improved gave me many special moments, for which I am very grateful. It was for this reason that I titled my Presidential Address to the Library Association of the United Kingdom, "A Debtor to His Profession"—the most special moment of all.

Eugene Garfield

Memories of the 1957 Dorking Conference

In the spring of 1957 I temporarily shared an office in Washington, D.C., at Thomas Circle with my new partner Harry Brager. He wisely recommended changing the title of *Management's Documation Preview* to *Current Contents* (*CC*).[1]

A few weeks later I received an unexpected invitation to discuss *CC* and citation indexing (based on my papers in *Science*[2] and the *Journal of the Patent Office Society*[3]) at an International Conference on Classification for Information Retrieval to be held in Dorking, England. The conference was only a few weeks off, and I was nearly broke. Using TWA's installment plan, I bought a round-trip ticket for $489.

Dorking was my first personal encounter with British information scientists. (I would meet J. D. Bernal in connection with the science of science movement but not until the following year in Washington.) It is quite possible that Jesse Shera, editor of *American Documentation*, had suggested I be invited. Other participants included Robert Fairthorne, D. J. Foskett, Eric J. Coates, Cyril Cleverdon, Brian C. Vickery, D. J. Campbell, N. T. Ball, Jack Wells, Barbara Kyle, John Mills, and last, but not least, S. R. Ranganathan.[4] I spoke at length with Ranganathan and others about my earlier meeting with Henry Evelyn Bliss in 1954.[5] The Bliss classification was better known in the United Kingdom than in the United States thanks to Jack Mills. FID publication #7146 commemorated the fortieth anniversary of Dorking. Unfortunately, I was unaware of this 1997 meeting, so I missed the opportunity to catch up with old friends, many of whom I had not seen for years. In that reminiscence of the Dorking conference Robert Fairthorne mentioned my "surprise" at the British members "disagreeing without being disagreeable,"[6] unlike the rancor frequently encountered at the early meetings of the American Documentation Institute. On the other hand, Cyril Cleverdon recalled the evening when "Gene Garfield defended his proposals for a citation index against a group of very skeptical and outspoken critics,"[7] including Cyril himself![8] Jean Aitchison recalled me as "a young man vigorously marketing his ideas of journal contents lists, at an extra evening session."[9] Indeed, 1957 was the year that the Life Sciences edition of *Current Contents* was introduced to the pharmaceutical industry.

The 1957 proceedings volume, on page 98, contains a concise account of citation indexes covered in the evening session on 14 May.[10]

On the second day I realized that if I attended Wednesday's session I would not see London. So I took an early morning train to Victoria Station. During the next fifteen hours I visited everything from the Tower of London to Parliament to the British Museum Library. I arrived at Victoria Station about midnight and was shocked to learn that it was closed. The only transportation to Dorking was a taxi. When I chaired the morning session the next day, the audience gasped when I said that I had taxied from London. I didn't mention that it used most of my remaining cash. This remarkable meeting eventually led to my joining the U.K. Institute of Information Scientists (IIS), which in 1966 gave me an honorary fellowship. Through IIS, I met and became friends with researchers like John Martyn, Alan Gilchrist, Charles Oppenheim, and others too numerous to mention. Somewhat later I met Tony Cawkell,[11] who became ISI's man in London and then director of research.

References

Note that Eugene Garfield's *Essays of an Information Scientist,* 15 vols. (Philadelphia, ISI Press, 1977–1993) may be accessed in pdf form through the following site: http://garfield.library.upenn.edu/essays/html.

1. Garfield, E. (1967, February 21). *Current Contents*—ninth anniversary. *Current Contents*. Reprinted in *Essays of an Information Scientist,* Vol. 1. (1977). Philadelphia: ISI Press, pp. 12–13.
2. Garfield E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science, 122*(3159),108–111. Reprinted in *Essays of an Information Scientist,* Vol. 6. (1984). Philadelphia: ISI Press, pp. 468–471.
3. Garfield, E. (1957, August). Breaking the subject index barrier—a citation index for chemical patents. *Journal of the Patent Office Society, 39*(8):583–595. Reprinted in *Essays of an Information Scientist*, Vol. 6. Philadelphia: ISI Press, pp. 472–484 (1984).
4. Garfield, E. (1984, February 6). A tribute to S. R. Ranganathan, the father of Indian library science. Part 1. Life and works. *Current Contents* No. 6, pp. 5–12. Reprinted in *Essays of an Information Scientist,* Vol. 7. (1985). Philadelphia: ISI Press, pp. 37–44; and Garfield, E. (1984, February 13). A tribute to S. R. Ranganathan, the father of library science. Part 2. Contribution to Indian and international library science. *Current Contents* No. 7, pp. 3–7. Reprinted in *Essays of an Information Scientist,* Vol. 7. (1985). Philadelphia: ISI Press, pp. 45–49.
5. Garfield, E. (1975, April 14). The "other" immortal: A memorable day with Henry E. Bliss. *Current Contents* No. 15, pp. 7–8. Reprinted in *Essays of an Information Scientist,* Vol. 7. (1985). Philadelphia: ISI Press, pp. 37–44.
6. Gilchrist, A. (Ed.). (1997). FID Publication 714: From classification to "knowledge organization": Dorking revisited or "past is prelude." The Hague: FID General Secretariat, p. v.
7. Ibid, p. vi.
8. Cleverdon, C. (1964). Citation indexing. *Nature, 203*(4944), 446; and Cleverdon, C. (1964). Science Citation Index, an international interdisciplinary index to the literature of science. *Revue Internationale de la Documentacion, 31*(4), 161.
9. Gilchrist, p. vii.
10. Garfield, E. (1957). In Summary of discussions. Proceedings of the International Study Conference on Classification for Information Retrieval. London: ASLIB/New York: Pergamon, p. 98.
11. Garfield, E. (1971, October 6). A. E. Cawkell. Information detective—and ISI's man in the U.K. *Current Contents.* Reprinted in *Essays of an Information Scientist,* Vol. 1. (1977). Philadelphia: ISI Press, pp. 245–246.

Robert McMaster, Madeline Berry, Krista Perry, James Perry, and Iver Igelsrud (from left to right).

Madeline M. (Berry) Henderson

Prolific Abstractors

In the early to mid-1950s there were three innovative approaches to the task of managing science information resources. In one, Calvin Mooers called for the application of "descriptors" to documents and for coding by random numbers and superimposed punching on edge-notched cards. The random number codes and superimposed punches would cut down on false drops. Mortimer Taube proposed the use of unit terms, found in documents, to be recorded on individual cards, with document numbers listed on the appropriate term cards. Searches involved combining or coordinating terms to define desired subjects and matching common document numbers on the selected cards. James Perry and his team of Perry, Berry, and Kent believed that the meanings of index terms or subject headings needed to be made explicit and proposed expressing the semantic elements or "factors" of such terms. They also described use of brief "telegraphic" abstracts to express document contents.

Each of the three innovators documented their approaches: Mooers issued Zator Company Technical Bulletins, while Taube published five volumes of his *Studies in Coordinate Indexing*. But neither matched the output of Perry, Berry, and Kent. We appeared in *Chemical and Engineering News* and in *American Documentation;* we conducted conferences and edited their proceedings; we published books on machine literature searching and on the use of punched cards. We were so prolific that we inspired Si Newman to compose a limerick. (Simon Newman was a chemist at the U.S. Patent Office, active in early efforts to automate patent searching procedures.) Si wrote:

Perry, Berry, and Kent
Re-announce the self-same event.
Their abstracts in miniature
Cover the world's literature
Recently doubled by Perry, Berry, and Kent!

It was fun—an exciting time to be active in the field—and I am glad to have been part of it.

Courtesy American Society for Information Science.

Eugene B. Jackson

Involvement of Technical Staff in Classifying Internal Research Reports

My own experiences with three major governmental-industrial special library research report systems (some founded before 1915) indicate the importance of involving the parent organization's technical staff (engineers, scientists, attorneys, etc.) in all aspects of the development and use of these systems. The systems in these organizations were successful and highly valued by those who used them. Many former users often returned to visit—and to show others—the libraries long after they had left or transferred from the government agency or company. This was even true of such notables as General "Hap" Arnold and Colonel Charles Lindbergh. The stories of the success of these information systems are actually the stories of dedicated library staff and colleagues. These stories are told in more detail in the ERIC document cited below.

Hope Thomas was hired in 1919 as index and catalog clerk at the McCook Field Library of the Army Air Service in Dayton, Ohio, which later became Wright Field Reference Library. Thomas began work on a technical reports classification and indexing system and developed it into one of the best and largest of its kind in the military services. She eventually became librarian and received an official commendation for her work. When I began work there in 1946, I was soon aware of the ways in which users valued the system she had built over the years. (Later, this library became an integral unit of the pioneering efforts of Wright Field to establish a firm presence in dissemination of foreign-language research reports.)

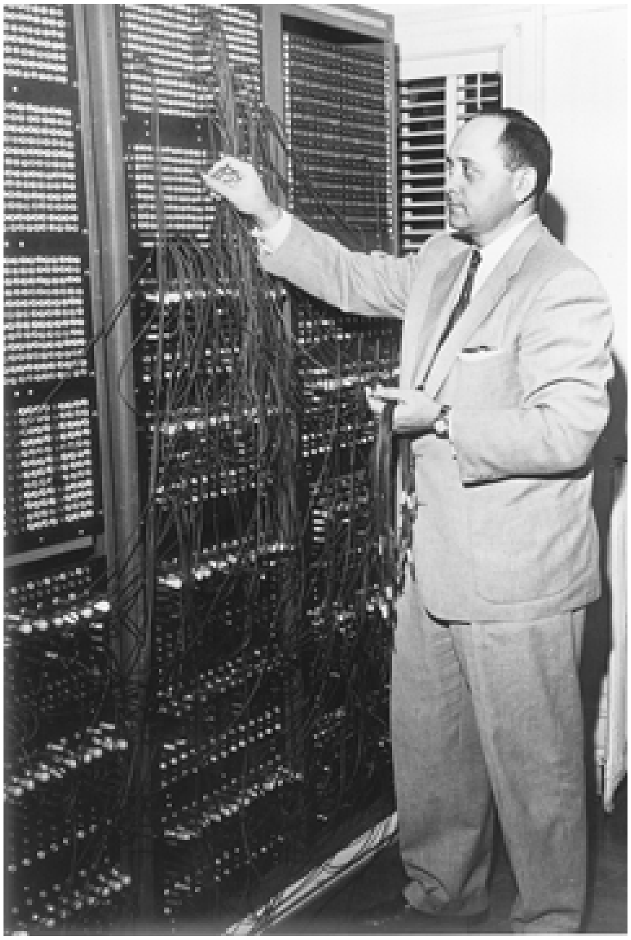
Caroline Lutz began her career as a clerk typist in 1917 at the National Cash Register (NCR) Company, also in Dayton. In 1925 she moved to Detroit, Michigan, to become the first librarian of the General Motors Research Corporation, where she rapidly developed a reputation for providing excellent service, particularly in the development of an internal technical reports system. When I became her successor in 1956, I was a direct beneficiary of her years of good work and heard many stories of how she had served her users. One of these grateful users, Charles Kettering, founder of Dayton Engineering Laboratories Company (DELCO) and her boss at GM since 1925, had turned over all the fees he received as an editor for *Readers' Digest* to the library. It was said that he told her to "buy good literature to civilize his engineers."

Emma Wedenbine also began her career as a clerk typist with the library of the NCR in 1922. She became the librarian for NCR in the 1920s and retired in 1972. She developed an outstanding reputation for service and for the development of a broad spectrum of special library activities, particularly in relation to internal technical reports. I knew her as a result of my work with the Cincinnati chapter of the Special Libraries Association. One of her accomplishments was the establishment of a regular branch of the Dayton Public Library adjacent to the NCR Engineering Library. It was also said that the founder of NCR, Colonel Deeds, wanted his engineers "civilized."

References

Jackson, E. B. (1978, July). "Three grande dames" or Dayton, Ohio, as a well spring of women special librarians. ERIC report, ED 184,503. Austin/Syracuse: ERIC Clearinghouse.

Jackson, E. B. (1986, Summer). Notes. Tracking the elusive special library for a quarter century. *Journal of Library History, 21*(3), 585–599.



Courtesy the Plain Dealer *(1958). All rights reserved.*

Allen Kent

Early in February 1958 my colleagues and I organized a national meeting at Western Reserve University in Cleveland to discuss a proposal to establish a national center for scientific and technical information. The stimulus for this proposal was the Soviet's launching of *Sputnik* in October 1957. Many U.S. scientists suggested that one of the reasons for the Soviets' taking the lead in the space race was the existence of their Institute of Scientific Information—which was characterized by a British scientist who visited as "really shattering . . . No other agency in the world is doing this job."

On 17 January 1958, two weeks before the meeting, an enterprising reporter for the *Cleveland Plain Dealer* learned of the forthcoming meeting and published a major story that occupied eight columns across the front page—"WRU Plans World Document Center."

The fallout from this story was dramatic: Newspapers, magazines, and wire services throughout the country picked up the story. And soon Senator Hubert Humphrey, then chairman of the Committee on Reorganization for the Senate Committee of Government Operations called us. Senator Humphrey opened the phone conversation with the question, "Just what are you fellows in Cleveland up to?"

The rest is history: Based on our phone discussion and subsequent visits, the senator organized hearings to evaluate the U.S. posture in my field of endeavor, which led to major funding for information programs from the National Science Foundation and other agencies.

In August 1958 I received a phone call from James Rand, president of Rand Development Corporation and chairman of President Eisenhower's Patent Council. He indicated that he was in New York City, shepherding a Soviet delegation headed by a minister of the U.S.S.R. The visit was occasioned by a letter from the Soviet premier to President Eisenhower requesting most-favored nation status in regard to financial credit. The request was denied, but it was suggested that an exchange of visitors to assess developments of mutual interest might lead to a reversal of the decision. Rand indicated that the visiting Soviet delegation had expressed considerable interest in my work and that an invitation for me to visit would be in order.

I met the delegation in New York, and it led to a month-long visit to the U.S.S.R. to explore the topic "Information Retrieval and Machine Translation—U.S. vs. U.S.S.R. Developments."

The chronicle of the visit is not the subject of this brief summary, but suffice to say, it was a turning point in my career and was the subject of an article in *Harpers Magazine* in 1959.

Fredrick G. Kilgour

Interlibrary Lending Online

Soon after World War II, various presidents of Ohio colleges and universities persuaded the Ohio College Association to seek means of making available to students and faculty of each institution the books and journals in the libraries of all Ohio institutions. In 1967, after years of hard work, the Ohio College Library Center, now the Online Computer Library Center (OCLC), was incorporated. Later that year I was appointed the executive officer, the first employee.

OCLC's long-term goals are

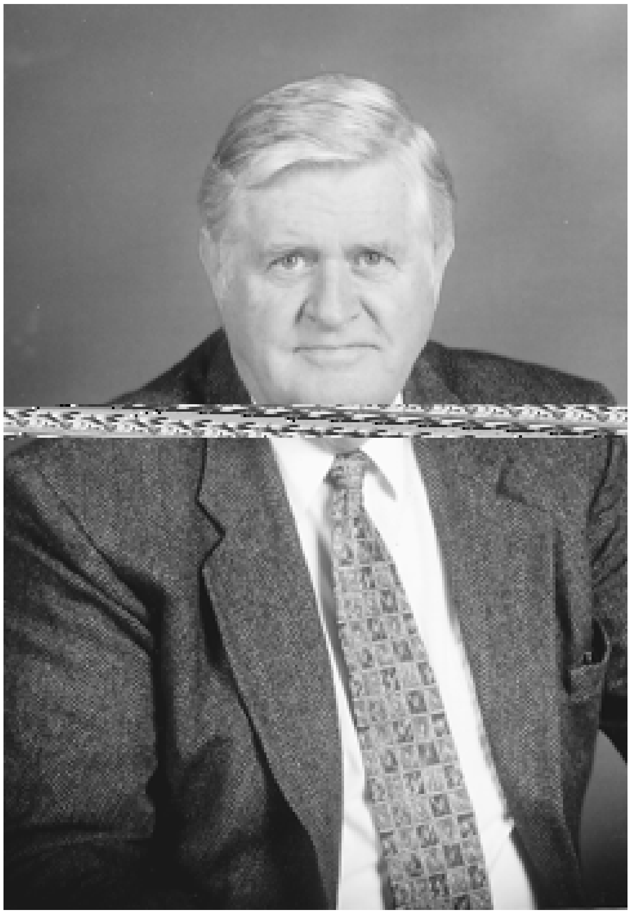
1. To increase availability of library resources to library patrons.
2. To lower the rate of rise of library per-unit costs.

To achieve these goals, OCLC designed an online cataloging system that reduced cataloging costs and simultaneously produced an online union catalog that revealed the location of books and journals in participatory libraries. This system was activated on 26 August 1971. Interlibrary lending was carried out by mail.

Not surprisingly the new system boosted interlibrary lending by Ohio academic libraries. In 1977 a study confirmed this increase. At the time of the study OCLC's online catalog contained slightly more than three million entries, and approximately 1,200 libraries were participating. Thirty-seven Ohio academic libraries possessed adequate data to participate in the study; they furnished interlibrary lending counts for three years before activation, 1968/69 to 1970/71, and for the following six years, 1971/72 to 1976/77.

According to the study, during 1968/69 to 1970/71, twenty-one libraries had lent a yearly average of 0 to 99; five had lent an average of 100 to 199; and eleven had lent 200 or more titles. Beginning with the third year after activation of the online system (1973/74), there was a sharp increase in interlibrary lending. The libraries experienced a growth in interlibrary loans that was 75 percent higher than it would have been had they not participated in the OCLC online system. The increase was especially dramatic for libraries that had done very little lending before joining the system. Percentage increase for the 0-to-99 group was 1,437; for the 100 to 199 group, 1,179; and for the 200-plus group, only 241. Five of the first group had never lent a title to another library during the initial three-year period, but by 1974/75 all twenty-one were participating in interlibrary lending.

After a trial period throughout the spring of 1978, OCLC activated its new online Interlibrary Loan Subsystem on 1 July 1978 and began charging for each request to borrow. The daily average for the first month was 1,488 transactions on 372,000 transactions a year. The 1997/98 *OCLC Annual Report* recorded 8.2 million online interlibrary loans transacted. In other words, OCLC has attained its first goal by increasing availability of library resources to 8.2 million library patrons.



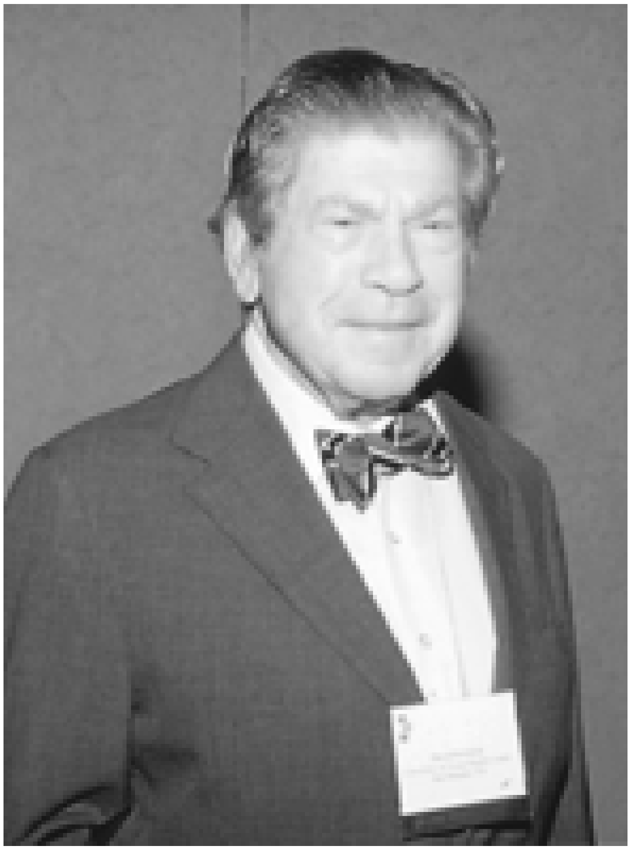
Donald W. King

Information Science Research

I came into information science through the back door. In 1961 Edward Bryant, James Daley, and I started a statistical consulting company, Westat Research Analysts, in Denver. After struggling for a year, we were awarded a contract with the U.S. Patent Office in Washington to help design experiments involving indexing and searching of patent documents. As part of this work I was asked to attend the 1962 annual meeting of the American Documentation Institute (predecessor to the American Society for Information Science). This was my introduction to information science and where I first met some of the attendees who were at the CHF Conference.

As a statistician I had the unique opportunity of observing a great deal of information science research from the early 1960s onward through statistical observation, statistical surveys, and experimental design. These studies were performed while I was with Westat, during a short sabbatical with Informatics, and at King Research (from 1976). They involved such interesting areas as evaluating information retrieval systems (U.S. Patent Office, CIRCOL, APA, NLM Cancerline) and information centers (NTIS, OSTI, DTIC, ERIC). In the 1970s my staff and I had the opportunity to describe scientific communication through a series of National Science Foundation studies under the umbrella "Statistical Indicators of STI Communication." During this time we also performed systems analysis and development of the International Cancer Research Databank (including development of Cancerline). We were also involved in editorial processing centers and electronic publishing and the development of a numeric metadata system for the U.S. Department of Energy. We performed two studies of the impact of the 1976 revision to the copyright law and a study of the cost-benefit of copyright formalities. I was also involved in studies of information professionals and information professional competencies. Beginning in the early 1980s Jose-Marie Griffiths and I developed a conceptual framework, methods, and measures for assessing the use, usefulness, and value of information services and products. We were able to apply our approach to well over one hundred libraries of all kinds and multi-type library networks across fourteen states and several regions. Since most of these studies were proprietary, I have spent much of my retirement aggregating and documenting these results with Jose so the knowledge will not be lost to others. Carol Tenopir and I are also documenting the results of over twenty thousand readership surveys and cost studies involving publishers, libraries, and scientists' time. Our intent is to help scientists, librarians, and publishers understand each others' participation in scholarly communication and how they should approach the transformation to electronic journals.

In addition to being able to describe and play a role in the growth of information science, my fondest memories come from my good fortune to have worked with so many talented people and to have observed the innovations and enormous contributions made by so many of the information pioneers. My greatest regret is that so much of their knowledge, contributions, and experience has been lost and is not being built upon by many of those involved today in digital libraries, electronic publishing, and other forms of scientific communication. I am hopeful that the effort to describe pioneering efforts in information science will bridge this "disconnect."



Courtesy Chemical Heritage Foundation.

David A. Kronick

Including History

It was my good fortune some fifty years ago to find a dissertation subject that has engaged me for over a half century. The subject was the origins and development of the scientific journal beginning with the scientific revolution in the seventeenth century, which provided the impetus for its origins, and ending with the chemical revolution at the end of the eighteenth century—a period in which journals had greatly proliferated. I did not know when I started how widely these journals were dispersed and that it would provide me with an incentive to visit great libraries when I had the opportunities to travel in Great Britain and Europe. In fact it led a friend to suggest that I was more motivated by an urge to travel than by a zeal for scholarship.

The subject, as you can imagine, has many ramifications: in the history of science, sociology, economics, and philosophy, which with the best intentions I was not able to explore adequately. It was enough, however, to learn that the issues that were relevant then—such as efforts to maintain the quality of the scientific literature, editorial methods and policy, determination of priority, secrecy, and disclosure—were much like the issues that engage us in scientific documentation today.

One of the conclusions I reached in my dissertation was that the scientific journal, as it was invented, fulfilled two distinct and different functions. First it served as a vehicle to disseminate information, and second it served as a depository, from which relevant items could be retrieved on demand. This finding, of course, did not startle anyone, because it was obvious that a single instrument that could serve both purposes was very efficient. It provided for a continuous flow of information, a means of quality control, and a location and citation capability. It could do so effectively, however, only if the necessary supporting secondary instruments that could provide the access to the depository were also developed. These kinds of instruments began to appear early in the history of the scientific journal and have been enlarged, refined, and modified ever since.

The technology for the development of the journal was in existence for over two hundred years, beginning with the invention of printing in the last half of the fifteenth century, before it was applied to the dissemination of scientific information. Today we are faced with a new technology that may have an equally important influence on the methods of disseminating and storing scientific information and, in fact, may be able to integrate the two functions and eliminate the necessity of waiting for periodicity in publication. This technology may not inaugurate the paperless society, which was being predicted a few years ago and which is easily refuted by anyone visiting a library photocopy room or computer search station, but it will provide new challenges and opportunities.

F. Wilfrid Lancaster

Getting Published

My biggest moment in the field of information science occurred in 1968 when I learned that my first book had been accepted for publication by John Wiley. Following work on the Aslib Cranfield Project and various evaluation studies for Herner & Company, I had recently completed a large-scale evaluation of MEDLARS. The book was based primarily on my experience in these various evaluation studies.

It is never easy for a relatively unknown author to find a publisher, and my experience was no exception. I first submitted the manuscript to Columbia University Press and later to McGraw-Hill. Both sat on the submission for several months before they eventually declined.

At that point I was ready to give up on the whole thing and was beginning to feel that the book was perhaps not worth publishing after all. Quite by chance I mentioned the situation to a professional colleague, Jesse Ostroff. He said that he was quite friendly with Joe Becker who, at that time, was very influential in Wiley's publishing in information science. Jesse gave a copy of my text to Joe, who liked it. John Wiley made a rapid decision to publish.

Since getting this book published made a significant contribution to furthering my career, I will be eternally grateful to Wiley, Becker, and, especially, Jesse. Jesse was an information specialist with some government agency, but I no longer remember which. Very fortunately for me, he had attended a workshop or short course I had given in the area of evaluation.

There are other people, of course, who profoundly influenced my career, and I owe them all a debt of gratitude: Cyril Cleverdon, who led me to the field of information retrieval; Saul Herner, who brought me back from England; and Herbert Goldhor, who offered me a full-time faculty position at Illinois and rewarded me with rapid promotions.

There have been many notable events in my career (which, incidentally, I do not regard as completely over), including many important awards from ASIS but getting my first book published was definitely the highlight.

J. Mills

Classification for Retrieval

We cannot remember the two most important events in our lives—our birth and our death. But I remember clearly the day that started me, a librarian, on the specialist's path of information classification for retrieval.

I had just been appointed as librarian of the City of London College. The college had been completely destroyed in the bombing raids of 1941 and had been reconstituted in a makeshift way in the vacated building of a large insurance company nearby. The library was a sad shadow of the fine one destroyed (which I had used as a student). My task was to build a new one, almost from scratch. My favorite professional aphorism has long been Jesse Shera's observation that two things distinguish the librarian's job: bibliography and retrieval. Bibliography stands for all the problems relating to the information-bearing materials themselves. Retrieval summarizes the central problem in the use of the materials (the information store), which is to find relevant items, in real time. Beginning from scratch the job of getting the library going brought these priorities home to me as nothing else could.

To classify the stock, I quickly realized, was the first priority if the stock was to be openly accessible to the students and staff. I began the job using the system with which I was most familiar—the Dewey Decimal classification. This system, a brilliant pioneer in its day, was quite inappropriate for what was needed now. The college was founded in 1848 (I began in its centenary year) expressly to meet the demand for professional studies in economics, finance and banking, insurance, accountancy, mercantile law, management, and other areas reflecting basic interests of the city. These closely related subjects are badly scattered in Dewey, and after a week or so of increasing exasperation, I decided to shut up shop and see if there was a better system available. So I went to the library of the Library Association, with its comprehensive collection of general and special classifications.

The scope of the college courses really called for a general system, and I soon decided that the new bibliographic classification of H. E. Bliss best fitted the bill. Although its final volume was still to be published, the H. W. Wilson Company proved very helpful and put me in touch with Bliss himself. Bliss then helped me resolve a number of problems posed by the crucial economics class—and I was on my way. I wrote about the problems of applying the bibliographic classification, was invited to lecture on classification and cataloging, and, in 1952, became a full-time lecturer in these subjects. At about the same time the Classification Research Group was formed, and I became a member at its inception, to my great benefit.

The emphasis in retrieval now is on the problems of micro-information. It is often assumed that the organization of knowledge in libraries is insufficiently important to warrant continued research. I think this is a very shortsighted view. A truly comprehensive, flexible, and logically structured map of modern knowledge, designed expressly to serve its central functions, is surely the least the library and information profession deserves now. Evidence that some librarians think so may be found in Cambridge. Here a number of the colleges have been reclassified by BC2. This is a new edition of Bliss's original system, with all its classes completely restructured, using the modern techniques of faceted classification and vastly enlarged. I am sure Dewey would approve!

Courtesy Chemical Heritage Foundation.

Herbert Ohlman

Mechanical Indexing Goes Public

It all came together for me in 1958 at the International Conference on Scientific Information (ICSI). I submitted a paper (the work for which had been done while I was employed by Carrier Corporation) to ICSI, "Subject-Word Letter Frequencies with Applications to Superimposed Coding." It was accepted by the chairman of Section 5, H. P. Luhn. Later, I received a set of preprints of all the conference papers.

Meanwhile, working with colleagues at the System Development Corporation (SDC), I had developed a system of mechanical indexing for SDC and Lincoln Laboratory documents. As a public demonstration of the capabilities of permutation indexing, I persuaded SDC to support an all-out effort to index the ICSI preprints in time for distribution at the conference. We distributed "A Permutation Index to the Preprints of the International Conference on Scientific Information" on the first day of the conference. At the same time Luhn of IBM and colleagues from their Service Bureau Corporation distributed a "Bibliography and Index: Literature on Information Retrieval and Machine Translation," which contained titles indexed by the keywords-in-context system—subsequently known better by its acronym, KWIC. We produced our index entirely with IBM tabulating (punched-card) machines, but Luhn used punched cards only for input, converted the data to punched-paper tape, and used a computer to produce the final index. However, the appearance of the printed indexes was practically identical.

Similar systems were developed during or just after World War II but this must have been the right moment for mechanical-computerized indexing to go public.

Claire Schultz

A Career Turning Point

One of the significant moments in my career was the day in 1953 when John Mauchly, co-inventor of the Univac computer, came to visit me at Merck, Sharp and Dohme. We had met a few days earlier, at the Welch Library Conference on information retrieval at Johns Hopkins University.

Mortimer Taube spoke at the evening session of the conference. He described his manual Uniterm system. I had not heard of it before and thought some of his statements ill founded: I said so when it was time for audience participation. Emcee, Ralph Shaw of the U.S. Department of Agriculture Library took me to task, saying that I did not show proper respect for Taube's insights as a logician. Mauchly responded in defense of what I had said and thereby became my hero.

Returning to Philadelphia by train, Mauchly chose to sit with me, to hear more about the Merck, Sharp and Dohme retrieval system to which I referred when I answered Taube. He knew Calvin Mooers, so understood the random, superimposed coding system I had described. He was interested in how we sorted our coded cards via the IBM 101 Electronic Statistical Machine, which had recently come on the market. At the end of the conversation Mauchly said he would visit my library (not far from where he lived), so he could experience its operation.

On his arrival he chose the subject for the search—treatment of anemia in human beings. We focused on the auxiliary panel board, which accommodated codes for up to four search terms per question, and allowed for *and*, *or*, and *not* connectives among them, without having to change any wiring of the machine between questions. Cards fed through the sorter were deposited in separate pockets of the machine, according to which descriptors the cards contained. At one point, Mauchly asked where I had learned Boolean algebra. I answered that I had not heard of Boolean algebra. He chuckled and said that I had absorbed it somehow and was using it to good advantage. He added that he had thirty people working for him who were solving numerical problems by computer. He needed someone who could work with language problems. Could I come work with him?

My response was that I did not know a thing about computers. He replied that he could teach me. He said he did not know much about information retrieval but that together we could do some very important work. My next response was that the system he was viewing was in its infancy and it would not survive if I left. He said he understood. Would I signal if I ever wanted to change jobs? Three years later I signaled, and he hired me.

Courtesy Chemical Heritage Foundation.

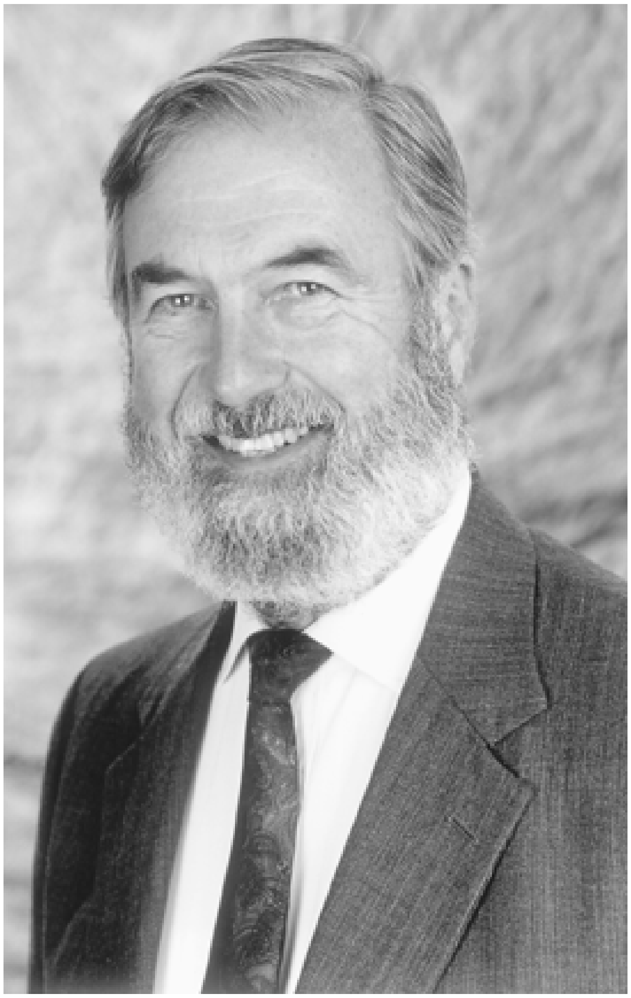
Winifred Sewell

Sixty Years of Progress

When asked to recall a memorable moment in two minutes, my reaction was that my most memorable moments are right now. I'm honored to be present with the giants of science information system creators at this particular point in time.

Today from my suburban Maryland home I'm working with an outfit in Syracuse, New York. They have a search engine that responds to queries in the user's own words with documents ranked according to their closeness to the question. To do so, their computer checks every word in all the documents in the database. This is what we dreamed of in 1961 when I was responsible for medical subject headings (MeSH) as a team member creating MEDLARS. At that time the inadequacies of our 64K computer and serial search techniques left us far short of what is being done today. However, as you are well aware, today's search engines aren't perfect either. There is a tremendous amount of work to be done, much of which requires the same thinking and challenges that we experienced in developing MeSH.

The fantastic difference today that boggles my mind is how much more productive we can be. Much of our research can be done electronically with findings pasted into our reports by a mouse click. Documents and messages can be sent by e-mail or fax. Compare these procedures with the clipping and pasting we did when studying subheadings in 1961 or with waiting three days in the 1940s to get photocopies of a document and taking three more to send it by snail mail. Viewing librarianship from the perspective of nearly sixty years is fun!



Courtesy Dialog Corporation. © Glenn Matsumara.

Roger Kent Summit

An Explorer of the Online World

As one of the earliest explorers in the online world—in fact, some have credited me with creating it—I have been asked to tell you a bit about the history of Dialog and how my original vision for the online industry is still relevant today.

It all began when I was a doctoral candidate at Stanford in 1960. I took a summer job at Lockheed Missiles and Space Company to improve information retrieval methods. Many of you are too young to remember the second-generation computers of that era. Suffice it to say, they used batch processing, which was cumbersome and required that you be a computer programmer to interact with a computer. Moreover, computers in those days were used mainly for accounting and scientific computation—not for processing text. The common argument around Lockheed was that it was usually easier, cheaper, and faster to redo scientific research than to find out if anyone had ever done it before!

There came a point when I got very excited about the possibility of using the computer for information retrieval. My feeling was that by using the computer we could make a significant contribution toward providing access to the world's published literature.

By the mid-1960s third-generation computers with random access disks, CRT terminals, and telecommunications ushered in the new possibility of interactive computing. A colleague and I proposed that Lockheed establish a lab to explore this new technology. Our primary goals for an information retrieval system were that

- It had to be command driven so that searchers could use it directly without needing computer programmers to act as intermediaries.
- It had to be recursive, meaning that there needed to be a means to limit or expand the hits from a search without having to re-enter the search.
- It had to provide an alphabetical display of all retrievable terms from which one could choose.
- It had to let searchers retrieve a few items at a time to see if their query was on target.

In 1968 we won our first major contract from NASA to develop an online retrieval system for their database of aerospace research documents. The result was NASA/RECON (Remote Console Information Retrieval Service), which permitted the searcher to enter several descriptors at once and get an immediate response. Furthermore, the search could be modified as you went along (i.e., recursion) without having to reenter the entire search. For example, engineers interested in an alloy's heat tolerance could enter the name of the alloy, the heat range or ranges that concerned them, and other relevant indexing terms. It sounds like ancient history today, and it is—but try entering that kind of search in one of today's popular Web search engines!

Subsequently, our group won contracts with the Atomic Energy Commission, the European Space and Research Organization, the U.S. Office of Education, the National Technical Information Service, and others to apply this retrieval technology to their databases.

Because interactive access proved of value to many organizations, in early 1972 we arranged to offer the ERIC (Educational Resources Information Center) and NTIS (National Technical Information Service) databases to any subscriber with a computer terminal. This is when the DIALOG Information Retrieval Service, named after its information retrieval language, became the world's first commercial online service.

Over the years the company has undergone changes in ownership and name, and Knight-Ridder Information continues to expand its products and services. But my dream that this company would be the primary source of access to professional information throughout the world has remained constant throughout its twenty-five-year history.

With the rapid growth of the Web, some have been predicting the demise of traditional online services. I don't agree. Recently, I was doing some research in preparation for a speech I presented in Stockholm. I determined that DIALOG contains more than twenty times the total amount of information accessible through the Web. Furthermore, the two have grown at roughly the same rate over the past year, based on AltaVista statistics.

In addition to comparing the quantity of information on DIALOG and the Web, I compared the quality of search results for several topics using DIALOG and the AltaVista search engine. I'm sure it will come as no surprise that the DIALOG results were highly relevant, while the AltaVista results were, to be generous, somewhat encyclopedic in nature. I found that it was difficult and often impossible to do a comprehensive and in-depth review of a particular topic on the Web.

It's somewhat ironic that with the phenomenal growth of the Web and concomitant advances in interface design, Web search engines lack even the most rudimentary features that were basic in the first online retrieval system we designed thirty years ago—such features as field specification, display of index terms, or options to allow one to refine a search.

Nevertheless, the Web has accomplished what the traditional online services have been unable to do before now—capture the interest of a broad base of end users.

Robert S. Taylor

For Whom We Design Systems

Probably the most significant moment in my professional career happened about 1953 when I was sitting at the reference desk in the Lehigh University Library. Note that this was well before the computer became a ubiquitous artifact in America. The computer at that time was a huge machine filled with tubes that required heavy air conditioning. There was no such thing as "online." The personal computer was thirty years away. I had been educated as a historian and had been at various times a newspaper reporter, sports editor, intelligence agent, freelance writer (unsuccessful), and now a librarian. I suddenly realized that in all my adult life I had been doing the same things: gathering, organizing, retrieving, analyzing, and communicating information. From this realization sprang a whole series of questions over the next several decades. What the hell am I doing? Is there a new grouping here (a new profession)? If so, how are such professionals educated? How do people seek and make use of information? How do we as professionals help people become aware of the significant role that information plays in their lives? How can we design systems that will help people resolve problems critical to them and, at the same time, enhance the quality of life around them? Over the next forty years I tried to find some satisfactory answers to these questions.

The profession and education: Through the Center for the Information Sciences at Lehigh, which I directed (1962–1967), the Program in Language and Communication at Hampshire College (1967–1972), and especially at the School of Information Studies at Syracuse University, both as dean and professor (1972—1983), I began to outline an education for this new profession.

Information systems: My work on value-added processes (NSF-supported), published in 1986, began to answer my system design questions.

Information seeking and use: Papers published on question negotiation and information seeking in libraries (1968) and on information use environments (1990) began to open up for me those concerns.

These questions are, I feel, fundamental to the profession and will remain so. Worth noting is that I place people at the center of my concerns. It is people, both as individuals and as members of organizations, for whom we design systems. This is a user-driven approach. Technology, important and overwhelming as it is at this moment, is but a means of gathering, storing, manipulating, and moving information to people who can make use of it. Our professional responsibility is to understand the technologies and to use them effectively to help people in whatever setting. Without people at the center we become but another technology-driven vocation.

Brian Campbell Vickery

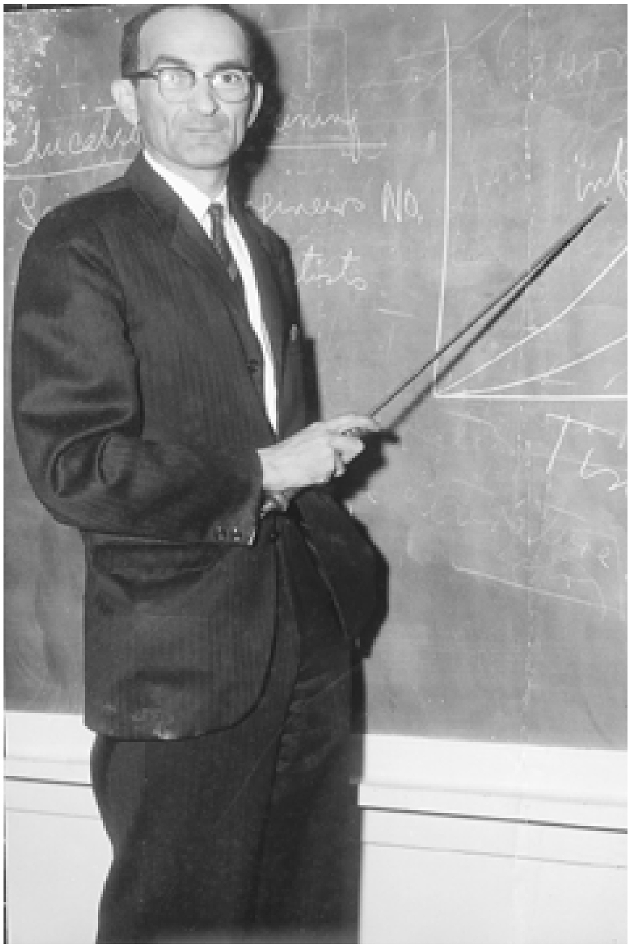
New Information Vistas

I would pinpoint 1958 as a special time in my career. I had for some years been working with the Classification Research Group in London, and in 1957 we had held a small but successful conference to which Jesse Shera, Gene Garfield, and others had come from the United States. In 1958 I published my first book, *Classification and Indexing in Science*, and attended the International Conference on Scientific Information in Washington. This was my first visit to the United States—I flew in a U.S. Army transport plane with Cyril Cleverdon. The conference papers opened up all kinds of new information vistas—in many ways setting the agenda for the ensuing development of the field. I met many interesting people—some who stand out in the memory are Peter Luhn, Mortimer Taube, John O'Connor, and Desmond Bernal. The experience of attending the conference, and of other visits I paid at that time, led to the writing of my second book, *On Retrieval System Theory*. During the conference Donald Urquhart told me that he would be looking for a deputy for the developing U.K. National Lending Library (NLLST), might I be interested? Yes, please! Despite all the excitement of online bibliographic access, in the last resort the provision of actual documents to working scientists is the end aim of scientific information provision, and in the United Kingdom we have always regarded the development of our national science library service as a major achievement.

I had taken a chemistry degree in 1941, and till I joined Urquhart in 1960, I worked as a chemist and then chemical librarian—first at a government explosives factory, then at Imperial Chemical Industries. Much of my work has been concerned with scientific and technical information. During the 1950s I started to compile a history of scientific communication, a work only recently completed, which I hope will soon be published. After leaving the NLLST, I carried out research, development, and consultancy for Aslib, and I particularly treasure our contributions to the development of the computer information systems of the Commonwealth Agricultural Bureaux and the House of Commons, work which led me to a third book on information systems.

I then moved to teaching and research at University College London. Information science defined as the practice of information provision has made enormous strides during the half-century I have been working in the field. But practice needs an underpinning of theory, and I have ever tried to explore and contribute to the development of information science in this second sense, summing up an understanding of it in the book written with my wife, *Information Science in Theory and Practice*. After retiring from full-time employment, we both had the opportunity of further active work developing online search aids (*Journal of Documentation*, June 1993).

In 1945 Desmond Bernal delivered an inspiring paper to an Aslib conference, asserting that "information service is essential to the progress of science." I am happy to have been able to make some contribution to its development.



Isaac D. Welt

A Brief Autobiography

I left Montreal, Quebec, where I had grown up, and, as a newlywed, enrolled in the Yale Graduate School, where I received a Ph.D. in biochemistry. After several years of published basic research in intermediary metabolism and an associate professorship at Baylor University Medical School in Houston, I moved to Washington, D.C., to enter the then-new field of information science at the Chemical-Biological Coordination Center, where using IBM cards and a primitive IBM mainframe, we attempted to correlate the chemical structure of numerous candidate drugs with their biological activity. This possibility was the main reason for my choice of information science as my future career in which I was active for some thirty-eight years as a university professor and research scientist.

As a result of a long-term, liberal research grant from the National Heart Institute, I organized and directed the Cardiovascular Literature Project, whose goal it was to collect and exhaustively *index* the effects of chemical substances on the cardiovascular system of humans and experimental animals. An M.D. colleague and I then set up a nonprofit organization called the Institute for the Advancement of Medical Communication, which unfortunately lasted only a few years.

I became quite active in ASIS and attended most of its conventions. Since I had ample travel funds, I also attended meetings of the Division of Chemical Documentation, the Medical Library Association, and the Special Libraries Association. I was one of the founders of the Drug Information Association, which has grown into a very large society. These organizations were instrumental in teaching me some of the facets of information science, since I never had a formal course in the subject. For a number of years I served as the associate editor of our *Journal of the American Society for Information Science*, which was also a valuable learning experience.

In 1964, after a few years of teaching as an adjunct member of the faculty, I was appointed to a tenured position as a full professor in the Center for Technology and Administration of American University in Washington, D.C. This group was a national pioneer in the teaching of computer usage to prospective managers at the master's degree level and was where I picked up enough knowledge about computer science (again, without formal course work). Later this group became the nucleus of American University's Department of Computer Science. I was most fortunate in being permitted to develop an entire curriculum in what could be termed "Computers and Scientific and Technical Information." I am most proud of the ten mature, adult practitioners whom I guided to their successful Ph.Ds.

I retired in 1992 as professor emeritus of scientific and technical information, which, to the best of my knowledge, is a truly unique designation.

Herbert S. White

Technology and Trivia

During the course of my professional career I have had the opportunity to work directly with three of the most amazing and innovative pioneers this profession has known: Hans Peter Luhn, Mortimer Taube, and Eugene Garfield. Any of the three could be the source of reminiscences by me of how the development of this young field was profoundly affected. However, as instructed, I will limit myself to only one such instance, and it concerns Hans Peter Luhn.

In the early 1960s, when we both worked for IBM, Luhn was concerned that promising new ideas were not being distributed throughout the corporation as rapidly and as fully as they should be. This was in part because of communications delays, in part because innovators did not necessarily know whom to tell, and in large part because bureaucratic managers would inhibit dissemination, perhaps, to protect their own managerial credit. Luhn wanted to establish a channel through which professionals could communicate their ideas immediately and directly to other professionals, without worrying about whether some individuals were being reached who perhaps did not have a need to know. Luhn's idea, of course, was not new. It was exactly the same premise as that developed by such individuals as Vannevar Bush, George Kistiakowski, and Karl Compton to speed up scientific and engineering communication during World War II. During wartime, waiting for formal communications patterns could have been disastrous. Luhn's contribution was in the use of the still young but nevertheless operational IBM computer system. Corporate management, whom Luhn could easily reach because of his many earlier accomplishments, agreed readily, and his idea was implemented over the objections of cadres of bureaucrats who wanted chain of command.

Luhn's idea worked, but he could not or at least did not foresee that individuals could and would misuse this new informal communications channel by using it for self-enhancement and self-publicity and by posting frequent (even daily) messages of "accomplishments" even when they had nothing to communicate. Eventually the system fell into disuse because really productive innovators could not risk exposing their time to trivial and self-serving inquiry.

I am brought to mind of the fact that we presently face a more modern version of this same dilemma. E-mail and list-servs allow us to communicate meaningful information rapidly and to a large audience. They also allow us to communicate trivia and garbage in the same manner. The limitation in all of this is that while technology has progressed rapidly, people have remained pretty much the same. It is for that reason that I, in my active and busy retirement, have limited myself to communication access that requires an individual effort. That means snail mail, telephone (I have no call waiting or message system—if there is no answer try again later), and fax machine. All of these require individual effort. I trust the technology; I just don't necessarily trust the people who use it.

Magda Whitrow

The Completion of the *Isis Cumulative Bibliography* (1923–65)

The high point of my career in the field of information science was the moment when, in the spring of 1984, I received from the publishers Mansell the last volume, Volume 6, of the *Isis Cumulative Bibliography* (1923–65), which I had been editing. The bibliography was based on references originally published in *Isis*, a journal for the history of science, which first appeared in 1913, edited by the distinguished historian of science, George Sarton. These critical bibliographies were compiled from information or offprints supplied by colleagues from all over the world.

When *Isis* was approaching the fiftieth year of its existence, the History of Science Society, which had taken over responsibility for its publication, received a small grant toward the production of an index. The editor, Harry Woolf, asked me to examine the problems involved and to make recommendations. There seemed to be a general consensus of opinion that the critical bibliographies formed an essential tool of research and that a subject index to these was most urgently needed. I suggested that these should be cumulated and republished in a fully classified form together with the necessary subject indexes. Following informal discussions on my report an editorial committee was set up by the History of Science Society under the chairmanship of I. B. Cohen (George Sarton's successor in the Harvard chair) and a substantial grant obtained from the National Science Foundation toward the production of a cumulative bibliography. It was Professor Cohen who suggested that, as a first stage in producing the cumulative bibliography, we should publish a volume containing all entries relating to the great men of science and other personalities of importance to the history of science, a kind of bio-bibliography. The project having been approved by the editorial committee and the Council of the History of Science Society, work began in 1944. The total number of personalities was in the region of ten thousand. The part of the bibliography dealing with personalities was published in two volumes (Volumes l and 2) and includes a section dealing with institutions. They were produced by a photolithographic process developed by Mansell for the 262 volumes of the *British Museum General Catalogue of Printed Books*. Advice was sought from experts on the possibility of using a computer-based process, but they decided that at the time this was not feasible.

Thanks to the generosity of learned societies, institutions, trusts, industrial concerns and private individuals from the United States and Great Britain it was possible to extend the project to four further volumes. Volume 3 covers the general history of science and the special sciences unrestricted by period or civilization, including special aspects of science and scientific disciplines. For classifying the material, I devised a scheme based on that used by Sarton. The schedules developed for the different subject fields are not very detailed, but they are faceted. The notation uses capital letters for subjects and lower-case letters for aspects. Volume 4 includes all entries referring to the early periods, including the Middle Ages, to Asian cultures, except the Near East, to African and American cultures. Volume 5 contains all those that refer to the modern period from the fifteenth to the nineteenth centuries. The project was completed by the publication of an author index.

Index