

Encoded Archival Description ¹

Die Entwicklung eines Kodierungsstandards für

Archivarische Suchhilfen

von DANIEL V. PITTI², Charlottesville (Virginia)

Abriß: Die *Encoded Archival Description (EAD)* nähert sich gegenwärtig der Vollen-
dung und der formalen Veröffentlichung als ein Standardverfahren. *EAD* versucht,
Hindernisse beim Erkennungszugang (engl.: intellectual access) auf geographisch
verteilte Primärquellen zu überwinden, indem es eine Standardkodierung für archiva-
rische Suchhilfen anbietet. *EAD* ist die z. Zt. aktuellste in einer ganzen Reihe von
Bemühungen, das Problem eines universellen Erkennungszugangs zu lösen. Wie
seine Vorgänger geht auch *EAD* dieses unter Berücksichtigung neuester technologi-
scher Entwicklung an. Die Technologie, die *EAD* zugrunde liegt, ist die *Standard Ge-
neralized Markup Language (SGML)* sowie die *Extensible Markup Language (XML)*.
Archivare sowie Größere und kleinere Archive sind in die Entwicklung von *EAD* mit-
einbezogen.

Encoded Archival Description

Da sich die *Encoded Archival Description (EAD)* ihrer Vollendung und formalen
Veröffentlichung als ein Standardverfahren nähert, erscheint es sinnvoll, sich des
langwierigen Problems zu erinnern, das *EAD* nun anzugehen sucht, außerdem die
verwendete Technologie zu untersuchen und sich den Prozeß zu vergegenwärtigen,
der Wesen und Struktur von *EAD* bestimmt hat.

Erfolgreiche Innovation geschieht nicht in einem Vakuum. Ihre intellektuelle Inspi-
ration erhält sie vielmehr aus der Tradition, auch wenn Innovation zur selben Zeit
eben jene zu verändern sucht. Unser Hauptanliegen bei der Entwicklung von *EAD*
war es, ein Instrument zu liefern, das helfen soll, die Beschränkungen zu verringern,

¹ Übersetzt aus dem Amerikanischen von Johannes Pfändner, Göttingen.

² **Über den Verfasser:** Daniel V. Pitti ist Project Director am Institut für fortschrittliche Technologie in
den Geisteswissenschaften an der Universität von Virginia. Zuvor war er der *Librarian for Advanced
Technologies Projects* an der Bibliothek der University of California, Berkeley. Pitti ist der Hauptent-
wickler der *Encoded Archival Description*.

die sich aus der geographischen Verteilung von Sammlungen für Forschende, Lehrende und andere ergeben und somit Suche und Benutzung von Primärquellen zu erleichtern. Moderne Versuche, Hindernisse zu überwinden, die aus der geographischen Verteilung von Quellen entstehen, datieren mindestens bis in die Mitte des neunzehnten Jahrhunderts zurück³. Bibliotheken und Archive haben seitdem immer wieder versucht, dem Problem mit den verschiedensten Mitteln zu Leibe zu rücken. *EAD* ist dabei zwar die neuste aber sicherlich nicht letzte Anstrengung in dieser fortschreitenden Tradition.

Versuche, das Problem der geographischen Quellenverteilung anzugehen, haben sich darauf konzentriert, universellen Erkennungszugang zu liefern. Bemühungen, das Problem des Zugangs auf die Materialien selbst bzw. ihren geistigen Inhalt zu lösen, stecken noch in den Kinderschuhen, da die technischen Möglichkeiten dafür erst seit kurzer Zeit zur Verfügung stehen. Als wir begannen, an einer computergestützten Datenstruktur für Suchhilfen zu arbeiten -- textanalytische Guides, die archivarische Sammlungen kontrollieren und beschreiben --, glaubten wir, daß ein solcher Standard ein wichtiger Beitrag zu dem lange ersehnten Durchbruch zum universellen Erkennungszugang sei, der außerdem dem Zugang auf die Materialien und ihren geistigen Inhalt selbst den Boden bereiten würde.

Es existiert eine enge Verbindung zwischen den Anstrengungen, die geographische Verteilung zu überwinden, und technologischen Innovationen. Alle Bemühungen, die in dieser Tradition stehen, folgen und sind inspiriert von neusten, vielversprechenden technischen Entwicklungen⁴. In dieser Hinsicht unterscheidet sich *EAD* nicht von seinen Vorgängern. Neuste Computer Hard- und Software hat in Verbindung mit Standardisierungsfortschritten und verbesserten Netzwerken die Phantasie all derer angeregt, die mit der Entwicklung von *EAD* befaßt sind.

³ Charles C. JEWETT: *Smithsonian Report on the Construction of Catalogues of Libraries and Their Publication by Means of Separate, Stereotyped Titles* (Washington: The Smithsonian Institution, 1853).

⁴ Die technische Entwicklung, die hinter Jewetts Plan stand, einen universellen Katalog zu entwickeln, war das Druckplattenverfahren. Anstelle von Metallplatten wollte Jewett allerdings Ton verwenden. Als der Plan dann fehlschlug, wurde er spöttisch als "Jewetts Matschkatalog" bezeichnet.

Die Entwicklung eines Standards ist nicht nur ein intellektuell-technologisches Unterfangen, sondern auch eine politische Übung; es handelt sich um eine Tätigkeit, die eine Gemeinschaft hervorbringt und definiert. Ein erfolgreicher Standard muß die Interessen einer Gemeinschaft widerspiegeln. Diese Interessengemeinschaft muß in seine Entwicklung involviert sein, wenn er ihren Interessen dienen soll. Daher haben wir bei der Entwicklung von *EAD* von Beginn an die Gemeinschaft der Archivare und Bibliothekare mit einbezogen.

Universeller Zugang über gedruckte Kataloge

Mit dem Versuch, die geographische Verteilung von Primärquellen zu überwinden, befindet sich *EAD* genau im Mainstream einer größeren Entwicklung des Bibliotheks- und Archivwesens, welche ihren Weg durch große Teile unseres Jahrhunderts unaufhaltsam gegangen ist. Schon lange bevor internationale Computernetzwerke und Online-Kataloge ihren Siegeszug antraten, stellte sich die Gemeinschaft der Bibliothekare zielbewußt den Herausforderungen der geographischen Verteilung von Sammlungen. Ursprünglich zielten diese Versuche darauf ab, landesweiten Zugang zu veröffentlichtem Material zu ermöglichen. Im Jahre 1909 traf die Kongreßbibliothek eine Vereinbarung zum Austausch von Katalogkarten mit mehreren größeren Bibliotheken. Herbert Putnam, damals *Librarian of Congress*, beschreibt den Plan und seine Absicht wie folgt:

Die *Library of Congress* wird jedem großen Forschungszentrum der Vereinigten Staaten die Kopie jeder Karteikarte, die sie für ihre eigenen Kataloge druckt, zur Verfügung stellen; diese werden dort eine Aufstellung darüber bilden, was Bestand der *National Library* ist. Wir hoffen, daß wir die Kopie einer jeden Karteikarte der *New York Public Library*, der *Boston Library*, der *Harvard University Library*, der *John Crerar Library* und verschiedener anderer Bibliotheken erhalten. Diese werden wir in einem Karteikartenkatalog großer Sammlungen außerhalb Washingtons ordnen und bewahren⁵.

Dies war der erste vorläufige Schritt zu dem, was später der *National Union Cata-*

⁵ The National Union Catalog Pre-56 Imprints (London: Mansell, 1968), vol. I, vii.

log werden sollte. Andere Bibliotheken schlossen sich an. 1926 hatte die Kongreßbibliothek nahezu zwei Millionen Karteikarten in einem Register zusammengetragen. 1948 wurde dieses Register offiziell "*National Union Catalog*" (*NUC*) getauft, und solche Bibliotheken, die bis dahin nur von ausgewählten Neuerwerbungen Mitteilung gemacht hatten, wurden nun gebeten, dies von allen Anschaffungen zu tun.

Die Titel zu sammeln war nur der Anfang der Bemühungen, eine gebrauchsfähige, landesweite Katalogisierung zu erstellen. Diese mußte, um universell gebrauchsfähig zu sein, universell zugänglich gemacht werden. Die Kongreßbibliothek brauchte bis 1956, um für dieses Problem eine Lösung zu entwickeln, indem sie nämlich den Buchkatalog wieder einführte, ein Format, welches die meisten Bibliotheken seit fünfzig Jahren nicht mehr verwendet hatten. 1946 veröffentlichte die Kongreßbibliothek "*A Catalog of Books Represented by Library of Congress Printed Cards Issued to July 31, 1942*". Zehn Jahre später wendete die *Library of Congress* auf Drängen der *American Library Association* diesen Zugang auch auf den *National Union Catalog* an und begann jene Titel in Buchform herauszugeben, welche die angeschlossenen Bibliotheken erworben hatten. Dies führte schließlich zur Veröffentlichung von mehr als 600 Bänden des "*National Union Catalog, Pre-1956 Imprints*", der größten Einzelpublikation, die jemals veröffentlicht wurde⁶. Zum ersten Mal hatten das Bibliothekswesen und die Öffentlichkeit, der es diente, ein System, um einen landesweiten nationalen Katalog zu erstellen und ihn generell verfügbar zu machen. Allerdings gewährleistete dieser landesweite Katalog nur Zugang zu veröffentlichtem Material und nicht zu den reichhaltigen Primärquellen des Landes.

1951 begann die *National Historical Publications and Records Commission (NHPRC)*⁷ damit, ein landesweites Register von Archiv- und Manuskriptsammlungen, die von den Magazinen des Landes unterhalten wurden, zusammenzustellen. Ziel war es, zentralen Zugang zu den Primärquellen des Landes zu ermöglichen. Zunächst konzentrierten sich die Bemühungen auf zusammenfassende Titelaufnahmen ganzer Sammlungen und nicht auf den Einzelzugriff auf Titel innerhalb von Sammlungskarteien. Auf diese Weise sammelte die *NHPRC* in den Fünfzigern Daten aus

⁶ The National Union Catalog Pre-56 Imprints, vol. I, x.

landesweit 1300 Magazinen, um dann 1961 "*A Guide to Archives and Manuscripts in the United States*" zu veröffentlichen⁸. Sie entschied sich 1974 zunächst, dieses Verzeichnis noch einmal zu überarbeiten und zu ergänzen, mußte aber bei näherem Hinsehen erkennen, daß die Anzahl von Archiven und Aufzeichnungen in den dreizehn vorangegangenen Jahren dramatisch gestiegen war, so daß sich zusammenfassende Titelaufnahmen ganzer Sammlungen aus Kostengründen von selbst verboten. Die *Commission* entschied sich, den Blick auf Informationen auf Magazinebene zu richten und somit einen größeren Zugang zu gewährleisten. Trotz dieses Blickwechsels behielt die *Commission* eine "*national collection-level data base on archives and manuscripts*" immer im Sinn⁹. Aus den verschiedensten Gründen wurde die Idee im Jahre 1982 aber dann verworfen.

1951, dem Jahr, in dem die *NHPRC* das *Directory* zu planen begann, startete auch die Kongreßbibliothek die Planungen zum "*National Union Catalog of Manuscript Collections*" (*NUCMC*)¹⁰. Der *NUCMC* sollte für Handschriften und Handschriftensammlungen das sein, was der *NUC* für Druckwerke war. Winston Tabb von der Kongreßbibliothek beschreibt ein Hauptargument für die Entscheidung, den *NUCMC* zu entwickeln:

Wissenschaftler, insbesondere im Fach amerikanische Geschichte, waren sehr hilfreich, die Gründung eines Zentrums anzuschieben, welches die Manuskriptsammlungsbestände, die der Forschung zugänglich sind, lokalisieren, aufzeichnen und publizieren sollte. Sie hatten lange gegen Schwierigkeit ankämpfen müssen, wenn sie bestimmte Handschriften lokalisieren oder auch nur Sammlungen identifizieren wollten, die möglicherweise Primärquellenmaterial enthielten¹¹.

⁷ Die *NHPRC* hieß zu dieser Zeit *National Historical Publications Commission*

⁸ Diese Zusammenfassung basiert auf Richard A. Nobles Artikel "The *NHPRC* Data Base Project. Building the 'Interstate Highway System'", *American Archivists* 51 (Winter 1988): S. 98-105.

⁹ ebd. S. 99

¹⁰ Diese Zusammenfassung beruht auf dem Vorwort zum *Library of Congress National Union Catalog of Manuscript Collections: Catalog 1991* (Washington, D.C.: Cataloging Distribution Service, Library of Congress, 1993), S. vii-ix.

¹¹ *Library of Congress National Union Catalog of Manuscript Collections: Catalog 1991*, S. vii.

Es dauerte bis Ende 1958, ehe die Kongreßbibliothek ihre Pläne mit Mitteln des *Council on Library Resources* in die Tat umsetzen konnte. 1959 wurde die Manuskriptabteilung innerhalb der *Library's Descriptive Cataloging Division* gegründet, und man beauftragte sie damit, das *NUCMC*-Programm einzurichten und zu betreuen. Der landesweite Handschriftenkatalog enthielt Titelaufnahmen auf Sammlungsebene über Sammlungen in Magazinen der Vereinigten Staaten, außerdem Einzelwerk-Titelaufnahmen für besonders wichtige Handschriften. Wie der *NUC* bestand der Katalog aus Karteikarten, und er sollte in Buchform veröffentlicht werden, erhältlich auf Abonnement. Der erste Band des *NUCMC* erschien 1962, ein Jahr nach "*A Guide to Archives and Manuscripts in the United States*" der *NHPRC*. Nach 32 erfolgreichen Jahren kündigte die Kongreßbibliothek 1994 an, daß Band 29 die letzte gedruckte Veröffentlichung des *NUCMC* sein werde.

Das Ende des *NUCMC* in gedruckter Form bedeutet nun keineswegs, daß es nicht mehr wichtig ist, landesweite Kataloge zu erstellen, die den Zugang zu unseren geistigen und kulturellen Ressourcen gewährleisten. Dieser Schritt war im Gegenteil die logische und vorausschauende Antwort auf die Erkenntnis, daß den Zielen des *NUCMC* besser gedient ist, wenn man sich leistungsfähige Computernetzwerke zunutze macht, anstatt bei der alten Drucktechnik zu verharren.

Universeller Zugang über Online-Kataloge

Zusammen mit dem Erscheinen maschinenlesbarer Katalogverzeichnisse lieferte die Einführung national vernetzter Computerdatenbanken den Archiv- und Bibliotheksgemeinschaften die Möglichkeit, zentralisierte, landesweite Kataloge zu erstellen, die jederzeit und überall zugänglich waren. Damit wurde auch der Boden bereitet, auf dem die Entwicklung standardisierter Verfahren wie *EAD* vonstatten gehen konnte. Zum ersten mal ermöglichte es die Technologie Archiven und Bibliotheken, einen Universalzugang zu gewährleisten, der nicht räumlich oder zeitlich begrenzt und somit viel effektiver und zugänglicher war als gedruckte Kataloge. Auch hat die Technologie das Anlegen landesweiter Datenbanken weitgehend vereinfacht. Indem sie Millionen maschinenlesbarer Katalogaufzeichnungen speicherten, entstanden aus den *OCLC*- und *RLG*-Datenbanken in den achtziger und neunziger Jahren de

facto landesweite Kataloge, und das nicht nur für die nationalen bibliographischen Einrichtungen, sondern zu einem Gutteil auch für den Rest der Welt. Wissenschaftler, Lehrende, die allgemeine Öffentlichkeit - wer auch immer netzwerkverbundene Computer im Büro, daheim oder in Bibliotheken verwendete, konnte erkunden, was an veröffentlichtem Material vorhanden war und wo man es finden konnte.

Bis 1993 enthielten die Aufzeichnungen in diesen nationalen Einrichtungen fast ausschließlich Angaben über Druckveröffentlichungen. Primärquellen in den Archiven des Landes und Handschriftenmagazine waren nicht erfaßt. Das sollte sich alles mit der Arbeit der *National Information Systems Task Force (NISTF)* der *Society of American Archivists* ändern. Von 1981 bis 1984 ebnete die *NISTF* geistig und politisch den Weg für die Entwicklung des *USMARC Archive and Manuscript Control (AMC)* Formates¹². Das *AMC*-Format versetzte Archive und Handschriftenmagazine in die Lage, kurze, zusammenfassende Surrogate für in ihrer befindliche Sammlungen im Rahmen von bibliographischen Katalogen anzubieten. Jedoch spezifizierte das *AMC*-Format selbst nur inhaltliche Kodierungsstandards. Es bot keine Standards für den eigentlichen Inhalt der Aufzeichnungen selbst. Ohne diese Standards war das Format aber nur eine leere Hülle. Die Gemeinschaft derer, die mit Archiven und Handschriften befaßt sind, befand die *Anglo-American Cataloging Rules, Second edition (AACR2)* für unzureichend. Das *AACR2*-Kapitel über die Katalogisierung von Handschriften verstieß gegen langjährige Prinzipien der Titelaufnahme. Steven L. Hensen, der zu diesem Zeitpunkt für die Kongreßbibliothek arbeitete, entwickelte daraufhin ein alternatives Regelwerk, das die Kodierungsstandards ergänzen sollte. Dieses Regelwerk, welches in seiner zweiten Ausgabe den Titel *Archives, Personal Papers, and Manuscripts (APPM)* trägt, ermöglichte es der Gemeinschaft von Archivaren und Handschriftenforschern in Verbindung mit dem *USMARC AMC*-Format, mehr als 475.000 Einträge auf Sammlungsebene zur *RLIN*-Datenbank der *Research Libraries Group* beizusteuern¹³. Durch diese Einrichtung haben Wissenschaftler nun Zugang zu einer wachsenden Ansammlung von Kurzbeschreibungen der Archiv- und Hand-

¹² Für eine kurze Geschichte und Bewertung der Arbeit der *NISTF* siehe David BEARMAN: *Towards National Information Systems for Archives and Manuscript Repositories. The National Information System Task Force (NISTF) Papers 1981-1984* (Chicago: The Society of American Archivists, 1987).

¹³ Die *RLIN* Statistiken wurden von Ann Van Camp von der *Research Libraries Group* zur Verfügung gestellt und und spiegeln die *RLIN* Datenbank, Stand 25. August 1997, wieder.

schriftensammlungen des Landes.

So wichtig und revolutionär diese Leistungen auch waren, so bedeuteten sie doch nur einen großen Schritt dahin, es Forschern zu ermöglichen, leicht an Primärquellenmaterial zu gelangen. Die verallgemeinerten Beschreibungen der AMC-Einträge können einen Forscher nur zu einer Sammlung führen, die für ihn relevante Quellen enthalten *könnte*. Der Forscher muß sich nun all den Bestandsverzeichnissen, Registern, Indices und Führern, sog. Suchhilfen, zuwenden, mit denen Büchereien und Archive administrativen und intellektuellen Zugang in Form von gründlichen und detaillierten Beschreibungen erlauben. Suchhilfen liefern hierarchisch strukturierte Darstellungen, die in abgegrenzten Schritten vom Allgemeinen zum Speziellen führen. Auf der allgemeinsten Stufe enthalten sie Informationen im Rahmen eines Katalogs auf Sammlungsebene. Auf der speziellsten Stufe benennen sie kurz einzelne Quellen. Dazwischen beschreiben sie in unterschiedlicher Genauigkeit untergeordnete oder Serien von verwandten Quellen. Suchhilfen sind detaillierte Landkarten, die von einer Hauptautobahn zu den Nebenstraßen führen und von da aus schließlich zum Ziel, der Quelle selbst.

USMARC AMC-Einträge auf Sammlungsebene und Suchhilfen sollen als Teile eines hierarchisch-archivarischen Zugangs- und Orientierungsmodells zusammenarbeiten. An der Spitze des Modells repräsentiert der *AMC*-Eintrag im Online-Katalog eine Sammlung und führt, über den Verweis "*mögliche Suchhilfen*", zur genaueren Information in der Suchhilfe. Im Gegenzug führt die Suchhilfe zu Materialien in der Sammlung.

In diesem dreirangigen Modell basiert die Titelinformation des Eintrags auf Sammlungsebene auf der Suchhilfe der Sammlung und wird auch von ihr abgeleitet. Nur ein kleiner Teil der Informationen aus den Registern und Bestandsverzeichnissen findet den Weg in den bibliographischen Eintrag. Die zusammenfassende Natur des Eintrags auf Sammlungsebene wird durch das Suchhilfen- und Katalogverzeichnis für die *National Municipal League Records 1890-1991 (bulk 1929-1988)* in der *Auraria Library* in Denver, Colorado dramatisch illustriert. Die Suchhilfe enthält mehr

als 1.400 Seiten und 30.000 Eigennamen. Im Vergleich dazu ist der *AMC*-Eintrag für diese Sammlung ungefähr zwei Seiten lang und hat neun Eigennamen als Zugangspunkte.

So positiv eine Entwicklung wie die des *AMC*-Formates auch war, so war sie doch noch nicht der letzte Schritt im Bemühen nach universellem Zugang zu Primärquellen. Dennoch war *AMC* ein ausgezeichneter Prolog zum letzten Akt. Die *AMC*-Einträge steigerten unseren Appetit nach mehr Informationen und ließen uns fast augenblicklich erkennen, wo wir nach ihnen suchen mußten, nämlich in den detaillierten Bestandsverzeichnissen und Registern, aus denen die Katalogeinträge auf Sammlungsebene von vornherein stammten. Fast gleichzeitig mit dem Triumph von *AMC* wurde klar, welches der nächste logische Schritt sein mußte, um Wissenschaftlern den einfachen Zugang zu relevanten Primärquellen zu erleichtern, ohne daß sie sich ein Flugticket kaufen oder die Vollendung ihres Forschungsprojekts dem Belieben der U.S.-Post anvertrauen mußten: Ein weiterer Codierungsstandard mußte her, einer, der den *AMC*-Standard komplettierte, ein Standard für die Suchhilfen selbst. Und es war genauso klar, daß dieser Standard landesweiten Internetzugang zu den amerikanischen Suchhilfen für Archive und Handschriften ermöglichen würde

Der Wert von Standards

Aber warum sollte man auf der Entwicklung eines Standards bestehen? Der Erfolg von *MARC* selbst sollte eigentlich jede Diskussion über die Notwendigkeit von Standards für die archivarische Gemeinschaft überflüssig machen. Neuste Erfahrungen haben jedoch gezeigt, daß uns der Reiz einfacher Techniken dazu verleiten kann, manches zu ignorieren, was wir gerade erst gelernt haben. In Zeiten knapper Budgets kann es schwierig werden, sich daran zu erinnern, daß wir die neuen Medien am besten nutzen, wenn wir sie mit bewährten Eigenschaften ausstatten. Es erscheint heutzutage schwierig, sich zu vergegenwärtigen, daß es wichtiger ist, unsere bewährten Methoden zu standardisieren, als uns kurzlebigen digitalen Moden in die Arme zu werfen, die nicht zukunftstauglich sind.

MARC hat erfolgreich Wert und Wichtigkeit eines allgemeingültigen Standards

demonstriert, indem er das Ziel erreichte, universellen Zugang zu Primärquellenmaterial zu ermöglichen. Wir arbeiten ständig und unermüdlich daran, übersichtlichen, universellen Erkennungszugang sowohl zu Primär- als auch Sekundärquellen zu realisieren. Diese bemerkenswerte Anstrengung wäre undenkbar ohne die Pionierarbeit der Bibliotheksgemeinschaft hinsichtlich der Entwicklung inhaltlicher und struktureller Standards. Mit der Entwicklung des *AMC* schloß sich die archivarische Gemeinschaft der Erkenntnis an, daß Standards eine übergeordnete Bedeutung besitzen. An die Vorteile von gut gestalteten, allgemeingültigen Titelaufnahmestandards gewöhnt, schien es unvorstellbar, daß die archivarische Gemeinschaft eigenständige, nicht standardisierte, oder noch schlimmer, unzulängliche Wege zum universellen Zugang zu den Suchhilfen und den Quellen, die sie dokumentieren, akzeptieren würde.

Standards sind die Grundlage, auf welcher Einzelne mit gemeinsamen Interessen Gemeinschaften bilden, in denen sie nicht nur koexistieren, sondern auch kooperativ gemeinsame und dauerhafte Arbeit leisten. Obwohl sich viele Archivare hinsichtlich der Adaption und Anwendung bibliozentrischer Bibliotheksstandards skeptisch zeigten, waren sie dennoch motiviert, mit der Bibliotheksgemeinschaft zusammenzuarbeiten, im gemeinsamen Bemühen, Archivmaterialien für die Benutzer leichter zugänglich zu machen. Archivare verfolgen wie Bibliothekare immer das Ziel, den Benutzern Informationen über Existenz, Zugänglichkeit, und Art der Materialien, die ihnen obliegen, so schnell wie möglich zur Verfügung zu stellen. Als daher die Instrumente zur Verfügung standen, um Surrogate für archivarische Sammlungen in öffentlichen Online Katalogen herzustellen, und es den Benutzern möglich wurde, relevante Primärquellen leichter zu finden und zu identifizieren, wurden diese Instrumente auch von der archivarischen Gemeinschaft angenommen.

Die Lehren aus dem *MARC*

Viele der Design-Eigenschaften, die so eindrucksvoll im *MARC* vorgestellt werden, wünscht man sich auch in einem Kodierungsstandard für Suchhilfen, und die Entwickler von *EAD* haben den *MARC* von Anfang an als Vorbild betrachtet.

Ein Kodierungsschema wie das des *MARC*, ein computerlesbares System also, welches die einzigartige intellektuelle Struktur von Katalogdaten repräsentierte, war absolut notwendig, wenn wir jemals große vernetzte Datenbanken bauen wollten, die weitgehende und effektive Kontrolle, Suche, Anzeige und Navigation innerhalb von Bibliothekssammlungen unterstützen sollten¹⁴. Der bloße Transfer komplexer Katalogeinträge in vernetzte Computer als unstrukturierter Text hätte die Rechner allein noch nicht in die Lage versetzt, Unterscheidungen und Beziehungen zu verwerten, aus denen deskriptiv katalogisierende Einträge bestehen. Katalogisieren ist eine "ordnende" Aktivität, die komplexe Regeln auf ein widerspenstiges, unregelmäßiges Informationsuniversum anwendet, um es in eine Form zu zwingen, so daß es den Benutzern geordnet erscheint¹⁵. Katalogmacher bestimmen die Identität von Autoren, Werken und ihren Themen und ihre Beziehung untereinander. Damit sie von Computern sinnvoll genutzt werden können, müssen alle diese ordnenden Dateien in einer Art und Weise vertreten sein, die es der Maschine erlaubt, sie mit beabsichtigten, vorhersagbaren Ergebnissen zu verarbeiten. Computer können keine verlässliche Datenverarbeitung mit flachen, unstrukturierten Texten leisten, weil die Programmierer ihnen nicht befehlen können, etwas zu verarbeiten, was sie vorher nicht identifiziert haben. Um vollen Nutzen aus der Computertechnologie zu ziehen, war es nötig, ein Kodierungssystem für Katalogdaten zu besitzen, welches die Grenzen seiner intellektuellen Bestandteile dem Programmierer und dem Computer gleichermaßen aufzeigte.

Für seine ursprünglichen Designer war *MARC* hauptsächlich eine Methode, die Herstellung gedruckter Katalogkarten zu automatisieren. Aber sie waren so weise, ein Codierungssystem einzuführen, das mehr als nur diesen einen Verwendungszweck erlaubte. Mit Blick auf die vielen späteren Verwendungszwecke für Katalogisierungsinformationen war es wichtig, daß sich das Kodierungsschema als ausreichend flexibel erwies, damit es allen möglichen Anforderungen auch gerecht wurde. Als der beste Weg, diesem Ziel gerecht zu werden, erwies es sich, das Schema deskriptiv und nicht prozedural zu gestalten.

¹⁴ "Kontrolle" meint hier zu "wissen, was es gibt und wo man es finden kann."

Prozedurale Kodierung sagt dem Computer, was er mit bestimmten Komponenten eines Textes zu tun hat; es liegt in seiner Natur, daß er die Information nur auf eine Art verwenden kann. Aber wir unterziehen, wie wir wissen, Katalogdaten einer ganzen Reihe von Verarbeitungsformen, um effektive Kontrolle, Suche, Anzeige und Navigation zu ermöglichen. Für die Anwendung auf multiple Datenverarbeitung müßte jede Komponente durch multiple Verarbeitungsbefehle kodiert sein. Dies ist jedoch in hohem Maße ineffizient, weil es viele Redundanzen mit einschließt. Unvorhergesehene Verarbeitungen schloß das Codierungsschema zum Zeitpunkt seiner Entwicklung außerdem aus.

Ein alternativer Ansatz ist es, die Informationen deskriptiv zu kodieren, und dieser ist von den Entwicklern des *MARC* klug gewählt. Deskriptive Kodierung bezieht die Kennzeichnung dessen mit ein, was eine wichtige Komponente ist: ein Katalogeintrag, ein Autor, ein Titel usw. Wenn wir wissen, was eine Datenkomponente ist, dann ist es möglich, verschiedene Verfahren auf sie anzuwenden, basierend auf der genauen Kenntnis ihres Wesens. Die Entscheidung, *MARC* zu einem deskriptiven Zeichensystem zu machen, stellte sicher, daß die Informationen in mehrfacher Hinsicht genutzt werden konnten. Es erlaubte außerdem unvorhergesehene Anwendungen in frühen Stadien der Entwicklung. Neben den zuverlässig repräsentativen Katalogdaten erkannten die Entwickler von *MARC*, daß ihr System einem allgemein gebräuchlichen entsprechen mußte, um sicherzustellen, daß die Kataloginformationen in einer sich ständig ändernden Rechnerumgebung auch Bestand haben würden. Ein Standard darf nicht an eine bestimmte Hard- oder Software gebunden sein, wenn er in einer ständig im Fluß befindlichen technischen Umgebung überleben will.

Das deskriptive Wesen der *MARC*-Codierung dient neben der Flexibilität der Datenverarbeitung auch dem langfristigen Überleben von *MARC*, indem es die Hilfsmittel bereitstellt, *MARC*-Daten in andere Rechnersysteme zu übertragen. Tatsächlich nutzen und speichern die meisten existierenden *MARC*-Systeme *MARC* nicht in seiner ursprünglichen Form; um dem Standard gerechtzuwerden, werden die *MARC*-Einträge einfach im- bzw. exportiert. *MARC* in einen zukünftigen Nachfolgestandard

¹⁵ David LEVY: Cataloging in the Digital Order, (<http://www.csdl.tamu.edu/DL95/papers/levy/levy.html>).

zu übertragen wird eine einfache Exportprozedur sein. Daß *MARC* sich in den unglaublich schnellen Veränderungen der Computerwelt der letzten 30 Jahre so beständig gezeigt hat, zeugt von der Weisheit seiner Entwickler. Diese Aspekte der Bauweise von *MARC* - die Tatsache, daß es sich um ein deskriptives Zeichensystem handelt und daß es von allgemeiner Gültigkeit ist - haben die Entwickler von *EAD* stark beeinflußt und das Wesen von *EAD* in hohem Maße bestimmt.

Bereits in einem frühen Entwicklungsstadium von *EAD* haben wir Optionen für die Codierung von Suchhilfen geprüft. Die primären Selektionskriterien waren zum einen, daß das gewählte System ein Standard sein mußte, welcher einer Reihe von formalen Konventionen des öffentlichen Gebrauchs entsprechen mußte und unabhängig von einzelnen Hard- oder Softwareproduzenten zu sein hatte. Zum anderen sollte er dem komplexen intellektuellen Gehalt und die Struktur von Suchhilfen zuverlässig entsprechen, und zwar so, daß hochspezialisiertes Suchen, Navigieren und Anzeigen möglich war.

Wegen der Designqualitäten von *MARC*, sein Erfolg im Erfassen des intellektuellen Gehalts bibliographischer Beschreibungen und der Umstand, daß es von Archivaren erfolgreich angewendet wurde, um den Sammlungen auch zusammenfassenden Zugang auf Sammlungsebene zu erlauben, wurde das System schnell als eine ernsthafte Option in Betracht gezogen. Der Standard war Allgemeingut geworden. Aber war er auch fähig, dem komplexen intellektuellen Inhalt und die Struktur von Suchhilfen zu entsprechen?

Nach gründlichem Studium und eingehender Überlegung gelangten wir zu der Erkenntnis, daß *MARC* aus drei prinzipiellen Gründen nicht das denkbar beste Schema war. Zunächst sind *MARC* Einträge auf eine maximale Länge von 100 000 Charakteren begrenzt. Dies entspricht ungefähr 30 8-1/2 X II Seiten von 10-pitch unformatiertem Text im ASCII. Weil viele Suchhilfen länger als das sind, verbietet diese Umfangsbegrenzung die Verwendung von *MARC*. Zweitens, und das ist noch wichtiger, paßt sich *MARC* hierarchisch strukturierten Informationen nur sehr schlecht an. Da

Suchhilfen per se hierarchische Dokumente sind, wird *MARC* durch seine flache Struktur zu einem unbefriedigenden Instrument. Wie Archivare aus schmerzlicher Erfahrung wissen, war *MARC* zunächst dazu bestimmt, Daten zu erfassen, die einen einzelnen bibliographischen Eintrag beschreiben. Komplexe Sammlungen, die abgestufte Ebenen der Analyse erfordern, überlasten die *MARC* Struktur sehr schnell. Meistens kann eine zweite Analyseebene hinzugefügt werden, aber die Art solcher Informationen ist begrenzt¹⁶. Der dritte Grund, *MARC* nicht für Suchhilfen zu verwenden, sind die Kosten. Es ist eine grobe Untertreibung, wenn man behauptet, daß Bibliotheken, Archive und Museen aus nicht gerade übervollen Kassen schöpfen können. Um es in Relation zu setzen: die Mittel für einen einzigen B-2 Bomber würden die *Library of Congress* mehr als drei Jahre finanzieren¹⁷. Aus finanziellen Gründen also konnte die Gemeinschaft der *MARC* Benutzer die Entwicklung modernster Hard- und Software nicht voranbringen.

SGML, HTML, XML UND EAD

Nachdem wir erkannt hatten, daß *MARC* keine angemessene Wiedergabe der Daten gewährleisten würde, wandten wir unsere Aufmerksamkeit der *Standard Generalized Markup Language (SGML)* zu. *SGML* liefert aus verschiedenen Gründen ein vielversprechendes Gerüst oder Modell für die Entwicklung eines Kodierungsschemas für Suchhilfen. Erstens ist die *SGML* genau wie *MARC* ein Standard (ISO 8879). Sie enthält eine Reihe von Konventionen des öffentlichen Gebrauchs und ist daher nicht Eigentum (und damit abhängig) eines bestimmten Hard- oder Softwareproduzenten. Zweitens kann sich *SGML*, anders als *MARC*, hierarchisch aufeinander bezogenen Informationen anpassen, und zwar auf so vielen Ebenen wie nötig. Drittens gibt es keine impliziten Umfangsbegrenzungen bei Dokumenten auf *SGML* Basis. Und schließlich ist der Markt für *SGML* viel größer als der für *MARC*.

¹⁶ Eine Möglichkeit, das Problem zu umgehen, ist es, multiple, hierarchisch aufeinander bezogene und miteinander verknüpfte Einträge auf verschiedenen Ebenen der Analyse zu verwenden: Sammlungsebene, Unterabschnitt und Einzelstück. Der Gebrauch multipler Einträge führt jedoch zu extrem schwierigen inter- und intrasystemischen Kontrollproblemen, die von den Entwicklern *MARC*-gestützter Software oder auch innerhalb des Formats nicht angemessen gelöst werden konnten. Auch wenn die Kontrollprobleme innerhalb des Systems befriedigend gelöst worden wären, hätte die Kontrolle zur hierarchischen Darstellung des multiplen Eintrags einen nicht mehr zu leistenden menschlichen Aufwand nach sich gezogen.

Obwohl *SGML* sowohl Standard als auch Allgemeingut ist, ist es keine *Markup Language* von der Stange, die man einfach mit nach Hause nehmen kann, um sie auf einen Brief, Roman, Artikel, einen Katalogeintrag oder eine Suchhilfe anzuwenden. Sie ist vielmehr ein Meta-Standard für *Markup Languages*, oder einfacher ausgedrückt, ein Standard um *Markup Languages* zu konstruieren. *SGML* liefert sowohl die formalen Mittel, um die logischen Komponenten oder Elemente von Dokumenten zu benennen, als auch eine Syntax und Meta-Sprache, welche die logische Struktur von Dokumenten und Beziehungen zwischen Dokumentenbestandteilen definiert. Sie ist ein Satz formaler Regeln zur Definition von spezifischen *Markup Languages* für individuelle Arten von Dokumenten. Wenn sie diese formalen Regeln anwenden, können die Mitglieder einer Gemeinschaft, die eine bestimmte Art von Dokumenten verwendet, zusammenarbeiten, um eine *Markup Language* zu schaffen, die sich speziell auf den gemeinsamen Dokumententyp bezieht.

Die spezifischen *Markup Languages*, die in Übereinstimmung mit den Erfordernissen der *SGML* geschrieben sind, heißen *Dokument Type Definitions (DTDs)*. Die *Assotiation of American Publishers* hat zum Beispiel drei *DTDs* entwickelt: eine für Bücher, eine für Zeitschriften und eine für Zeitschriftenartikel. Ein Konsortium von Softwareentwicklern und -Produzenten hat eine *DTD* für Computerhandbücher entwickelt. Die Kongreßbibliothek testet derzeit den Entwurf zu einer *USHARC DTD*. Die *Text Encoding Initiative* hat eine komplexe Reihe von *DTDs* zur Wiedergabe von literarischen und linguistischen Materialien entwickelt. *DTDs*, die von einer Gemeinschaft geteilt und befolgt werden, werden selbst zu Standards.

Während sich *MARC* der Strukturierung einer bestimmten Art von Daten widmet, nämlich Katalogdaten, ist *SGML* sehr allgemein und abstrakt. Es befindet sich formal oberhalb individueller *Markup Languages* für spezifische Dokumentklassen. Weil Syntax und Regeln der *SGML* formal und präzise sind, ist es möglich eine Software zu schreiben, die relativ leicht eingerichtet werden kann, um mit jeder kompatiblen

¹⁷ Laut HWW Seite der United States Air Force beträgt der Preis eines einzelnen B-2 Bombers 1,3 Milliarden \$. Mehrere andere Quellen beziffern ihn eher auf zwei Milliarden \$. 1997 betrug das Budget der Kongreßbibliothek 360.896.000 \$.

DTD zu arbeiten. Typischerweise besitzt ein *SGML* Softwareprodukt ein *Toolkit*, welches es dem Benutzer erlaubt, seine Funktionalität auf ein spezifisches *DTD* zu übertragen. Daraus resultierend schließt der Markt, der die *SGML* Softwareentwicklung vorantreibt, alle *SGML* Benutzer mit ein. Ähnlich dem *MARC* soll die *SGML* eher das deskriptive als das prozedurale Zeichensystem eines Textes begünstigen¹⁸. Wie oben bereits erwähnt, spezifiziert das prozedurale Zeichensystem (oder Markup) eine einzelne Prozedur, die auf eine Dokumentenkomponente angewendet wird, während das deskriptive Zeichensystem jede Komponente definiert und die Datenverarbeitungsroutine den Anwendungen überläßt.

Es ist sinnvoll das deskriptive Zeichensystem in zwei Arten zu unterteilen: in ein strukturelles und ein nominelles Markup. Das deskriptive strukturelle Markup identifiziert Dokumentenkomponenten und ihre logischen Beziehungen. Strukturelle Elemente sind generell Komponenten, die die deutliche, visuelle Präsentation bestätigen: z.B. Kapitelüberschriften, Paragraphen, Listen und Blockzitate. Das deskriptive nominelle Markup identifiziert Namenseinheiten, und zwar sowohl konkrete als auch abstrakte, wie z.B. Firmennamen, Personennamen, Themengebiete, Genres und Ortsnamen. Während ihre spezifische visuelle Präsentation sinnvoll erscheint, erfordern es solche Elemente üblicherweise, in einer Weise verzeichnet zu werden, die den Zugriff auf einige Aspekte des Dokuments erlaubt. Es ist auch möglich, *SGML* so zu verwenden, daß die deskriptiven Komponenten von Suchhilfen als Namenseinheiten verwendet werden. *EAD* unterscheidet beispielsweise Umfang und Inhalt, Biographien und Behördengeschichten, chronologische Aufzählungen, verschiedene Formen von Verwaltungsinformation und noch viele andere Komponenten der Archiv-Titelaufnahme. Weil die Software genau diese Komponenten erkennt, kann sie dazu benutzt werden, jede von ihnen in bestimmter Weise zu benennen, zu suchen, anzuzeigen und zu steuern.

Desweiteren unterstützt *SGML* das referentielle Markup. Wie der Name schon sagt, bezieht sich das referentielle Markup auf nicht vorhandene Informationen; es ist

¹⁸ Zur detaillierten Beschreibung der verschiedenen Markup Typen siehe James H. COOMBS, Allen H. RENEAR, und Steven J. DEROSE: Markup Systems and the Future Scholarly Text Processing. (Communications of the Association for Computing Machinery 30 (II), 1987, S. 933-947.

ein Markup in der dritten Person sozusagen. Referentielles Markup wird in der Regel für Hypertexte und Hypermedien verwendet, die die Grundlage für dynamische Verweise oder Links zu anderen Texten und zu digitalen Originalen oder digitalen Wiedergaben von Photographien, audio und audiovisuellen Materialien, Zeichnungen, Gemälden, dreidimensionalen Objekten aller Art, chemischen Formeln, Druckseiten, Musik, Choreographien und allem anderen bilden, was digital in sinnvoller Form aufgezeichnet und wiedergegeben werden kann. Zusätzlich zu seinen vielen anderen Vorteilen eröffnen uns Suchhilfen auf *SGML* Basis die aufregende Option, Zugang zu digitalen Wiedergaben von Primärquellen in unseren archivarisches Sammlungen zu ermöglichen.

Die *Hyper Text Markup Language (HTML)* ist eine *SGML DTD*, welche auf einen enormen Erfolg als unterstützender Codierungsstandard für das *World Wide Web* verweisen kann. Als eine spezifische *SGML* Anwendung beschränkt sich die *HTML DTD* auf einfache prozedurale Kodierungsvorgänge, die der Onlineanzeige und den Hypermedialinks dienen. Weil die *HTML* die Onlineanzeige von Suchhilfen vereinfacht und relativ leicht angewendet werden kann, gingen viele davon aus, daß sie zur Kodierung von Suchhilfen ausreichen würde. Wir hatten allerdings den starken Eindruck, daß sie wegen ihrer prozeduralen Fixierung dem komplexen intellektuellen Gehalt und der Struktur von Suchhilfen nicht ausreichend gewachsen war, um ausgedehntes Suchen, Steuern und Anzeigen zu ermöglichen. Beweise für die begrenzte Fähigkeit von *HTML*, intelligentes Suchen und das Erschließen von Dokumenten zu fördern, lassen sich schnell finden, ganz zu schweigen von komplexer Anzeige, Steuerung und weiterer Datenverarbeitung. Wer von uns hat sich nicht schon bei der Suche nach bereits bekannten oder themenverwandten Einzelheiten im *Web* der Hilfe von Suchmaschinen bedient und wurde von der Fülle von Informationen geradezu erschlagen. Allzu häufig wird unsere Geduld arg strapaziert, nur um dann ein oder zwei Dinge zu finden, die wir gerade benötigen¹⁹.

Der Erfolg der *HTML* als ein Anzeigeformat für das *Web* zeigt die eine wesentliche

¹⁹ Als Antwort auf dieses Problem hat die OCLC seit 1995 internationale Anstrengungen unternommen, um ein einfaches generisch-deskriptives Metadatenschema zu entwickeln, das es möglich machen würde, *HTML* Dokumente im *Web* intelligenter zu benennen und zu suchen.

Schwäche derzeit erhältlicher *SGML* Software überdeutlich, nämlich die im Moment beschränkten Möglichkeiten, die ursprüngliche *SGML* über das Internet zu verbreiten. Die Entwickler der *SGML* Software haben sehr gute und preiswerte *Tools* herausgebracht, mit denen man das Schreiben, Editieren, die Datenumwandlung und die Gliederung und Suche von Datenbanken in *SGML* bewältigen kann. Außerdem gibt es sehr gute *Publishing Tools* für hausinterne Veröffentlichungen und solche auf CD-Rom. *SGML* Dokumente über das *Web* zu verbreiten ist allerdings bis heute mit ernsthaften Hindernissen verbunden. Doch sind die Aussichten, daß sich das in nächster Zukunft ändert, gar nicht schlecht.

1996 gründete das *World Wide Web Consortium (W3C)* die *XML Working Group*, um mit der Aufstellung einer Reihe von Spezifikationen die Anwendung von *SGML* im *Web* zu erleichtern²⁰. Die *Working Group* schrieb in kurzer Zeit eine Spezifikation für ein vereinfachtes Unterprogramm von *SGML* mit dem Namen *Extensible Markup Language (XML)*. Sowohl Microsoft als auch Netscape haben sich verpflichtet, ihre Internet Browser mit *XML* auszurüsten.

Das Motiv, das die Entwicklung von *XML* vorantreibt, ist die Erkenntnis, daß *HTML* dem komplexen, gemeinschaftsorientierten Gebrauch der von vielen geteilten Informationen im Internet nicht gerecht wird. *HTML* verknüpft eine Reihe von prozedural orientierten */tags/*. Diese Reihe von */tags/* zusammenzubringen hat die Hestellung von Anwendungen vereinfacht, die das Leben von Autoren und Herausgebern im *Web* erleichtern. Seine einfache Benutzung ist ein wesentlicher Faktor für den Erfolg des *Web*. Die kleine in sich geschlossene */tag/-Reihe* hat jedoch ihren Preis: *HTML* hat eine extrem begrenzte Funktionalität. Jon Bosak hat drei Bereiche ausgemacht, in denen die *HTML* Mängel aufweist: Extensibilität, Struktur und Validation²¹. *SGML* zeigt sich in all diesen Bereichen leistungsstark, aber diese Stärke hat wie die Schwäche von *HTML* ihren Preis: Sowohl für Entwickler als auch Benutzer von Anwendungen erweist sich die *SGML* als kompliziert. Die *XML Working Group* der *W3C*

²⁰ Der Originalname war *SGML Editorial Review Board*. Jon Bosak von Sun Microsystems ist Vorsitzender der *Working Group*. Unter den Mitgliedern sind außerdem Jean Paoli, bei Micro Soff Entwicklungschef des Internet Explorer, und Tim Bray als Vertreter von Textuality und Netscape.

²¹ Jon BOSAK: XML, Java, and the future of the Web. (<http://sunsite.unc.edu/pub/sun-info/standards/xmi/why/xmiapps.htm>).

ging diese Schwäche damit an, daß sie einige *SGML Features*, die nur schwer zu installieren sind, benannte und untersagte. Das Resultat ihrer Arbeit ist *XML*, ein vereinfachtes Unterprogramm von *SGML* für die Benutzung im *Web*.

Die fortschreitende Entwicklung von *XML* und eng verwandten Standards verspricht die Überwindung des letzten Haupthindernisses bei der Verwendung von *SGML* für die Kodierung elektronischer Suchhilfen und ihrer leichten Verbreitung über das Internet²². Glücklicherweise wurden die meisten untersagten *XML Features* in der *EAD DTD* nicht verwendet, und diejenigen *EAD*-Ausdrücke, die untersagte *Features* verwenden, können leicht auf *XML* kompatible Weise ersetzt werden. Daher werden nur wenige Modifikationen nötig sein, um die zukünftigen Internetbrowser von MicroSoft, Netscape und anderen Anbietern zu nutzen, und diese Veränderungen werden noch vor der Veröffentlichung der *EAD* Version 1.0 abgeschlossen sein.

EAD als eine *SGML DTD* zu entwickeln scheint noch immer eine vorteilhafte Entscheidung gewesen zu sein. Sie erlaubt es uns, die Stärken von *MARC* zu übernehmen - deskriptives Markup und öffentliches Eigentum - und seine Schwächen zu überwinden - begrenzte Aufzeichnungs- und Speicherkapazität, unzureichende Hierarchie und geringe Marktattraktivität. Am Beginn der Entwicklung von *EAD* war es eine Glaubensfrage anzunehmen, daß das *Web*, um wirklich robust zu werden, der *HTML* entwachsen müßte und daß ihr wahrscheinliche Nachfolger auf der *SGML* beruhen würde. Dies war ein kalkuliertes Risiko, aber es scheint durch und durch gerechtfertigt gewesen zu sein.

Die Verankerung von *EAD* im Mainstream der Bemühungen von Bibliotheken und Archiven, zu universellem Zugang zu gelangen, und die Verwendung neuester Computer- und Netzwerktechnologien scheint *EAD* mit allem auszurüsten, was es braucht, um erfolgreich zu sein. Aber das erscheint nur dann, so wenn man das wichtigste Element bei der Entwicklung aller Standards ignoriert die Gemeinschaft.

²² Die *XML* enthält drei verwandte Initiativen: die *XML*, die *Extensible Linking Language (XLL)* und die *Extensible Stylesheet Language (XSL)*. Zu aktuellen Informationen über den Entwicklungsstand der jeweils letzten Entwürfe siehe <http://www.sil.org/sgmi/xmi.html>.

Überblick über die EAD Entwicklung

Der Erfolg eines jeden Standards hängt von der breiten Anteilnahme der Gemeinschaft an seiner Entwicklung und der weitgehenden Anerkennung seines Gebrauchswertes ab. Standards sind das Produkt von Gemeinschaften, und nicht das von Individuen, die in der *Splendid Isolation* arbeiten, und der Entwicklungsprozess ist eine Politische Übung und ein technisches und intellektuelles Unterfangen gleichermaßen. Ein Kodierungsstandard muß daher, um erfolgreich zu sein, die gemeinsamen Interessen der archivarischen Gemeinschaft und der Kräfte und Institutionen, die die Gemeinschaft unterstützen, reflektieren und voranbringen.

Seit Beginn der Entwicklung eines Kodierungsstandards für Suchhilfen erkannten diejenigen, die an den Bemühungen beteiligt waren, daß es von entscheidender Bedeutung sein würde, die archivarische Gemeinschaft in sein intellektuelles und technisches Design miteinzubeziehen. Als im Jahre 1993 die Mitarbeiter der *Berkeley Library* erstmals darüber nachdachten, einen solchen Standard zu entwickeln, betonten Jackie Dooley und Steven Hensen die Notwendigkeit, die Gemeinschaft auf breiter Basis miteinzubeziehen, wenn sich die Anstrengung lohnen sollte.

Das *Berkeley Finding Aid Projekt (BFAP)*, finanziert mit Mitteln aus dem *Title IIA Program* des Erziehungsministeriums, setzte einen Prozess in Gang, der zu *EAD* geführt hat. Es war das Ziel des *BFAP*, durch die Entwicklung eines *DTD* Entwurfs (ursprünglicher Name: "*FindAid*") und eine *DTD* gestützte, internetkompatible Datenbank zu demonstrieren, daß ein auf *SGML* basierender Standard sowohl durchführbar als auch wünschenswert war. Um sicherzustellen, daß der *DTD* Prototyp in Inhalt und Struktur auch den Suchhilfen der Gemeinschaft entsprach, sammelten die Mitarbeiter des *BFAP* repräsentative, qualitätsunabhängige Beispiele von Suchhilfen aus unzähligen Magazinen²³.

²³ Die Reaktion auf diese Sammlung erlaubt einen interessanten Blick auf den Entwicklungsprozess des Standards. Viele Magazine versprachen enthusiastisch, Suchhilfen zu schicken, aber nach einigen Wochen geduligen Wartens erreichten uns nur eine handvoll. Die Mitarbeiter des *BFAP* wandten sich an jedes Magazin, das eine Zusage gemacht hatte mit der erneuten Bitte, die Suchhilfen doch zuzusenden. Denn obwohl sie die Bemühungen von ganzem Herzen unterstützten, waren die Magazine doch besorgt darüber, wie die Kollegen ihre Suchhilfen wohl einschätzen würden. Die schließlich eingesandten Suchhilfen waren denn auch eher die, in die sie das größte Vertrauen setzten. Auf diese Weise begann die Gemeinschaft, sich freiwillig zu vereinheitlichen.

Anfang 1995 gingen die Eigentumsrechte an den Ergebnissen der *BFAP* an die landesweite Gemeinschaft über, wozu zwei Entwicklungen beitrugen. Im April sponsorten die *Comission on Preservation and Access (CPA)* und die *Library at the University of California, Berkeley* gemeinsam eine Suchhilfenkonferenz, bei der 70 Vertreter spezieller Sammlungen, Archive, Bibliotheken und Museen anwesend waren. Zweck der Konferenz war es, die Ergebnisse des *BFAP* zu bewerten und Empfehlungen abzugeben, was als nächstes getan werden könnte. Man stimmte enthusiastisch überein, daß das *BFAP* trotz seiner begrenzten Möglichkeiten ein großer Erfolg war und daß die Bemühungen weitergehen müßten, wenn auch unter der aktiven Teilnahme von Experten für Archiv-Titelaufnahme.

Die Möglichkeit, Archivexperten stärker in das Projekt einzubinden, kam zeitgleich mit der Bewerbung des Verfassers um eine Teammitgliedschaft im *Bentley Library Research Fellowship Program* der *University of Michigan*. In dem vom Verfasser geleiteten Team wirkten sowohl bekannte Archiv-Titelaufnahme Experten als auch versierte *SGML* Fachleute mit²⁴. Im Juli 1995 traf sich das Team in Ann Arbor, um das *BFAP* und sein *DTD* formal zu bewerten und um ein neues Modell zu entwerfen. Das Team gelangte zur schnellen Übereinstimmung über die Designprinzipien, die die *Ann Arbor Accords* genannt wurden, und verbrachte den Rest der Woche damit, ein neues Modell zu entwerfen, auf dem eine neue *DTD* basieren sollte²⁵. Dort wurde das *BFAP* auch in *Encoded Archival Description (EAD)* umbenannt.

Ein Wust von Aktivitäten folgte dem Ann Arbor Treffen. In den nächsten zwei Monaten schrieb der Verfasser den ersten Entwurf der *EAD DTD*. Während des Jahrestreffens der *SAA* im September 1995 begannen die Mitglieder des Teams damit, geeignete Mechanismen für die umfassende Einführung und Aufrechterhaltung eines Kodierungsstandards für Suchhilfen festzulegen. Die Designprinzipien und das überarbeitete Datenmodell wurden dem *SAA Committee on Archival Information Ex-*

²⁴ Weitere Mitglieder der Gruppe waren: Steven J. DeRose, Elec-tronic Book Technologies; Jackie Dooley, University of Cali-fornia, Irvine; Micheal J. Fox, Minnesota Historical Society; Steven Hensen, Duke University; Kris Kiesling, Unversity of Texas, Austin; Janice Ruth, Library of Congress; Sharon Gibbs Thibodeau, National Archives and Records Administration; Helena Zinkham, Library of Congress.

change vorgelegt, und das *Committee* wurde eingeladen, formal an der Entwicklung von *EAD* mitzuwirken. Das *CAIE* stimmte zu und gründete die *EAD Working Group (EADWG)*, deren Vorsitzender Kris Riesling war und in der Vertreter der *Library of Congress (LC)*, der *RLG*, des *OCLC* und der *SAA* mitwirkten. Das *CAIE* beauftragte die *EAD Working Group* damit,

1. die Entwicklung und Überarbeitung eines Datenmodells für Suchhilfen zu unterstützen,
2. die *EAD DTD* zu überarbeiten,
3. die *EAD DTD* zu testen und zu bewerten,
4. die Anwendungsrichtlinien zu überarbeiten, und
5. die Überarbeitung von *EAD* durch das *SAA Standards Board* und das *SAA Council* zu veranlassen.

Die *SAA* stimmte außerdem zu, das *LC Network Development/MARC Standards Office* förmlich zu ersuchen, die Wartung von *EAD* zu übernehmen, sobald dieses einer gründlichen Prüfung durch die Gemeinschaft unterzogen und als Standard akzeptiert worden war. Im Oktober 1995 sponsorte die *National Digital Library* der Kongreßbibliothek ein Treffen des Teams in Washington, D.C., um das Modell zu überarbeiten und einen *DTD* Entwurf vorzulegen.

Die *ATLIS Consulting Group*, die bei der Kongreßbibliothek unter Vertrag steht und mit der sich der Verfasser ständig berät, begann nach dem Treffen im Oktober mit der Bearbeitung der *DTD* und der Erstellung einer */tag/-Bibliothek*. In einem Brief an Susan Fox, *Executive Director* der *SAA*, stimmte das *ND/MSO* förmlich zu, als Wartungsinstanz für *EAD* zu fungieren, während die *SAA* die Verantwortung für die fortschreitende intellektuelle Aufsicht und Entwicklung des Standards übernahm.

Im Dezember 1995 erhielt die *SAA* vom *Council on Library Resources* Finanzmittel, um Anwendungsrichtlinien für *EAD* zu erstellen, und das *EAD* Team traf sich im

²⁵ "Ann ARBOR ACCORDS: Principles and Criteria for an SGML Document Type Declaration (DTD) for Finding Aids", *Archival Outlook* (January 1996), S. 12-13.

Rahmen einer Konferenz bei der *UCLA* vom 4. bis 6. Januar 1996 mit Anne Gilliland-Swetland und Thomas La Porte, um den *DTD* Entwurf und die */tag/*-Bibliothek zu überarbeiten und den Inhalt der Richtlinien zu skizzieren. Weitere Veränderungen wurden in die "*Alpha*"-Version der *DTD* eingearbeitet, welche zur ersten Installierung im Februar 1996 fertiggestellt und elektronisch veröffentlicht wurde. Vom 27. bis 29. April 1996 traf sich das *EAD* Team in Berkeley, um den Richtlinienentwurf von Gilliland-Swetland und LaPorte zu erörtern und um Änderungsvorschläge zu prüfen, die Teammitglieder und die ersten Installierer hinsichtlich der *DTD* "*Alpha*"-Version gemacht hatten. Die abgestimmten Veränderungen wurden in die "*Beta*"-Version eingearbeitet, welche vom Verfasser am 15. Juni 1996 abgeschlossen und nach Prüfung durch das Entwicklungsteam im darauffolgenden September öffentlich zugänglich gemacht wurde. Der Richtlinienentwurf, die */tag/*-Bibliothek sowie kodierte Beispiele einer großen Vielzahl von Suchhilfen wurden im Dezember 1996 der Öffentlichkeit über das Internet zur Verfügung gestellt.

Im Verlauf seines Entwicklungsprozesses begannen mehrere größere Forschungs- und Demonstrationsprojekte *EAD* zu installieren. Von Anfang an begannen die *UC Berkeley*, die *Duke University* und die *NDL* der Kongreßbibliothek damit, *EAD* für ihre Suchhilfen zu verwenden, um seine intellektuelle und technische Funktionsfähigkeit zu testen. Auch Harvard und Yale fingen an, gleich nach der Veröffentlichung Anfang 1996 mit der *Alpha-DTD* zu arbeiten. Die *University of California*, San Diego, begann erfolgreich damit zu experimentieren, wie sich datenbankgestützte Suchhilfen in *EAD* übertragen ließen. *SOLINET* entschloß sich, *EAD* in sein vom Handelsministerium finanziertes Monticello Projekt zu übernehmen, und auch die *NEH*-finanzierte *Dance Heritage Coalition* entschied sich, *EAD* für ihr Archivzugangsprojekt zu verwenden.

Seit die *Beta-Version* von *EAD* im September 1996 veröffentlicht wurde, haben mehrere Magazine Suchhilfenprojekte von unterschiedlicher Größe und Komplexität gestartet. Das *Public Record Office* in London arbeitet zur Zeit an einer Strategie zur Umstellung seines Magazinführers. Wenn dieser Führer fertiggestellt ist, wird er mehrere hunderttausend Seiten enthalten, die mehrere Jahrhunderte des britischen

Staatsarchivs umfassen. Mehrere britische Universitäten, darunter Liverpool, Oxford, Durham und Glasgow haben wichtige *EAD* Projekte auf den Weg gebracht. In den Vereinigten Staaten ist die *UC Berkeley* mit Finanzierung der *NEH* in das *California Heritage Digital Image Access* Projekt eingestiegen. Das Ziel dieses Projektes war es zu zeigen, daß *USHARC* Einträge auf Sammlungsebene in Verbindung mit *EAD*-verschlüsselten Suchhilfen einen effektiven und brauchbaren Zugang auf Sammlungen gewährleisten können, die mehr als 25.000 digitale Abbildungen von Bildmaterial aus der umfangreichen Bildersammlung der *Bancroft Library* umfassen, welche die kalifornische Geschichte und Kultur dokumentiert. Bezeichnenderweise wird der Prototyp des Zugangssystems für das *California Heritage Project* auf ein *Interactive University Project* genanntes K-12 Erweiterungsprogramm der Universität Berkeley angewandt, welches aus Mitteln des Handelsministeriums finanziert wird. An diesem Projekt arbeiten Fakultäts- und Bibliotheksangehörige mit K12 Lehrern und Lehrplannern aus den Schulbezirken von San Francisco und Oakland zusammen, um Lehrprogramme und Stundenpläne zu erstellen. Diese sollen das digitale Archiv nutzen, damit in den Klassen des Schuljahres 1997/98 (und möglicherweise darüberhinaus) Fächer unterrichtet werden können, die sich mit der Geschichte und den Kulturen Kaliforniens befassen.

Das *California Heritage Project* hat außerdem die Trägerschaft für zwei weitere Projekte übernommen, nämlich das *NEH*-finanzierte *American Heritage Virtual Archive Project* und das *University of California EAD Project (UCEAD)*, das vom Büro des *UC* Präsidenten als das erste in einer Reihe von *UC*-weiten digitalen Bibliotheksprojekten finanziert wird. Neben der Einrichtung einer *UC*-weiten Datenbank für Suchhilfen ist es eine Kernaufgabe der *UCEAD*, an allen neun Standorten Archive in die Lage zu versetzen, *EAD* durch die Benutzung gebräuchlicher Software *Toolkits* effizient installieren zu können. Das *American Heritage Project* beinhaltet die Zusammenarbeit zwischen den Universitäten von Stanford, Virginia, Duke und Berkeley; es ist das Ziel des Projektes zu zeigen, daß *EAD* einheitlich auf verschiedenartige existierende Suchhilfen für Sammlungen, die Erbe und Kultur Amerikas in den vier angeschlossenen Magazinen dokumentieren, angewendet werden kann und die Magazine in die Lage versetzt, ein gemeinsames virtuelles Archiv aufzubauen. Das

Projekt erforscht die intellektuellen, politischen und technischen Kernfragen, die geklärt werden müssen, um zusammenfassenden Zugang zu den Suchhilfen möglichst vieler Institutionen zu ermöglichen²⁶. Das Kernstück des Projektes ist die Entwicklung "eines annehmbaren Rahmens einheitlicher Handhabung" in der Anwendung von *EAD* auf bestehende Suchhilfen. Bei einem Treffen in Berkeley im November 1996 konnten die Vertreter der vier zusammenarbeitenden Institutionen, gestützt auf die weitreichende Arbeit von Berkeleys technischen und archivarischen Mitarbeitern, nach eingehender Debatte zu einem Konsens gelangen. Der erreichte Konsens wurde im ersten Entwurf der *EAD Retrospective Conversion Guidelines* festgelegt. Bald danach trafen sich Archivare der neun Standorte der Universität von Kalifornien in Los Angeles, um den Konsens der *Guidelines* weiter zu vertiefen. Die angeschlossenen Magazine hoffen, daß diese als Diskussionsgrundlage für einen landesweiten Konsens über "einen annehmbaren Rahmen einheitlicher Handhabung" dienen können, der den Weg zum universellen Zugang zu Primärquellen frei macht²⁷.

Die *Research Libraries Group* hat erkannt, daß die Entwicklung von *EAD* Ausbildung entscheidend für die Akzeptanz und Anwendung innerhalb der Gemeinschaft sein wird. Im Sommer 1996 entwickelte die *RLG* in Zusammenarbeit mit der Universität Berkeley das *Finding Aid SGML Training*, auch *FAST Workshop* genannt. Im Laufe des letzten Jahres hielt die *RLG* mit Mitteln der *Delmas Foundation* und des *Council on Library Resources* verschiedene Workshops in Großbritannien, Kanada und den Vereinigten Staaten ab. Unter der Agide von Michael Fox und Kris Kiesling konnten die *FAST Workshops* erfolgreiche Grundlagenausbildung für nicht wenige

²⁶ Unter den derzeitigen technischen Beschränkungen versucht das Projekt, die Suchhilfen in einer einzigen, zentralen Datenbank zusammenzufassen. Während sich die Technologie für den Zugang zu regional unterschiedlichen Datenbanken stetig verbessert, ein Modell, daß aus mehreren praktischen Gründen bevorzugt wird, werden die Lehren aus diesem Projekt den Übergang in die neue Technologie insprieren.

²⁷ Laut des Berichts von Richard Noble im 1980 erschienenen NHPRC Report schätzten die Mitarbeiter der Kommission, daß 20.000 Magazin- und über 700.000 Sammlungstiteleinträge in einer nationalen Datenbank enthalten sein würden. Vgl. dazu Richard A. Nobles Artikel "The NHPRC Data Base Projekt: Building the 'Interstate Highway System'" *American Archivists* 51 (Winter 1988): 100: Die Suchhilfen der Berkeley Datenbank haben einen durchschnittlichen Umfang von 27 Seiten, wenn dieser Durchschnitt repräsentativ ist, umfaßten 700.000 Suchhilfen fast 19 Millionen Seiten Text!! Es ist wert festgehalten zu werden, daß nach nur elf Jahren sich über 700.000 Einträge auf Sammlungsebene in der RLIN Datenbank befinden. Da viele der archivarischen Sammlungen des Landes niemals datenverarbeitet, geordnet und beschrieben worden sind, dürfte 700.000 noch eine geringe Schätzung sein.

Archivare leisten. *FAST* und andere Workshops haben dazu geführt, daß eine ganze Reihe anderer Magazine unterschiedlicher Größe inzwischen ihre eigenen Projekte für Suchhilfenkodierung begonnen haben. Die Universitäten von Iowa und Vermont, die *New York Public Library*, die *North Carolina State University* und die *University of North Carolina* haben, nur um ein paar Beispiele zu nennen, alle ihre Projekte auf den Weg gebracht. Im August 1997 übergab die *RLG* den Workshop der *Society of American Archivists* im Rahmen deren Jahrestreffens in Chicago. Die *SAA* hat nun den Workshop in ihre Ausbildungspläne eingegliedert.

Sowohl die *RLG* als auch Chadwyck-Healey erforschen Möglichkeiten, *EAD* in ihre Produkte und Dienstleistungen einzubauen. Im Zuge der erfolgreichen *EAD* Ausbildung hat die *RLG* eine Beratergruppe gegründet, die bei der Planung und Installation neuer Dienstleistungen behilflich sein soll. Zu diesem Zeitpunkt der Planung beabsichtigt die *RLG* landesweiten Zugang zu Suchhilfen in der ganzen Welt zu ermöglichen, und zwar sowohl zu denen, die auf örtlichen Servern untergebracht sind, wie auch zu solchen, die aus Mangel an Quellen oder aus dem Bestreben heraus, eigene Suchhilfen zusammenzustellen, von den Magazinen auf dem *RLG* Server abgelegt sind. Die Beratergruppe hat erkannt, daß die angeschlossenen Magazine *EAD* einheitlich anwenden müssen. Unter diesem Aspekt hat sie sich entschlossen, die vom *UCEAD* und den *American Heritage Projects* gemeinsam entwickelten *Retrospective Conversion Guidelines* dazu zu verwenden, Diskussionen anzustoßen, die zu gemeinschaftsweiten, "bestmöglichen" Richtlinien führen sollen. Darüberhinaus prüft die *RLG* die Möglichkeit, einen retrospektiven Umwandlungsservice zu betreiben, der Gebrauch von Drittanbietern machen würde. Chadwyck-Healey denkt derzeit über einen ähnlichen Service nach und sucht nach Möglichkeiten, sein *Archives USA Produkt* durch die Verwendung von *EAD*-verschlüsselten Suchhilfen zu verbessern. Neben den Aktivitäten von *RLG* und Chadwyck-Healey haben auch eine Reihe anderer Softwareanbieter *EAD* Produkte in Entwicklung.

Zusätzlich zu der erfolgreichen Übergabe des *FAST* Workshops an die *SAA* gab es im Rahmen des 97er *SAA* Jahrestreffens noch eine Reihe anderer wichtiger Entwicklungen. Jackie Dooley, Vorsitzender des Ausschusses für Veröffentlichungen,

berichtete von Gesprächen mit der Kongreßbibliothek, in denen es um die Veröffentlichung der *EAD DTD*, der */tag/-Bibliothek*, der Richtlinien und eines Praxisleitfadens ging. Kris Kiesling, Vorsitzender der *EADMG*, gab bekannt, daß sich Meg Sweet vom britischen *Public Record Office* der *Working Group* angeschlossen und die *Delmas Foundation* ein Treffen der *Working Group* im Herbst '97 in Washington, D.C. finanzierte. Bei diesem Treffen wird die Arbeitsgruppe die Vorschläge der internationalen Archivgemeinschaft prüfen und nach eingehender Diskussion entscheiden, welcher Verschlüsselung von *EAD* anlässlich der Veröffentlichung Anfang 1998 der Vorzug gegeben wird. Die *Working Group* wird außerdem die Entwürfe der */tag/-Bibliothek* und der Richtlinien, welche gemeinsam mit der *DTD* veröffentlicht werden, abschließend prüfen.

Schlußbetrachtung

Vor dem Erscheinen von *HARC AMC* und *APPM* hatte die Gemeinschaft der Archivare wenig Grund, deskriptive Standardverfahren zu entwickeln. Die ökonomischen Vorteile gemeinsamer Katalogisierung, die von den Bibliotheken genutzt werden konnten, standen den Archivaren nicht zur Verfügung, deren Sammlungen für gewöhnlich einzigartig sind. Trotzdem wollten sie ihr Material zugänglicher machen, ein professionelles Ziel, daß sie mit ihren Bibliothekskollegen teilten. Dieser Wunsch lieferte die Motivation, *MARC AMC* und *APPM* zunächst zu erforschen und schließlich zu vereinnahmen, mit dem Erfolg, daß die archivarische Gemeinschaft von Wert und Wichtigkeit der Kodierungs- und deskriptiven Inhaltsstandards überzeugt war. Außerdem waren die Archivare von dem Gedanken erfüllt, über die zusammenfassenden Titelaufnahmen hinauszugehen und einen Weg zu den vollständigen, detaillierten Suchhilfen zu finden, die den Kern aller Bemühungen bilden, archivarische Sammlungen zugänglich zu machen.

Das Auftauchen des Internets, das auf revolutionäre Weise die zeitlichen und räumlichen Grenzen unserer Informationsgesellschaft sprengte, weckte eine beständige aber schlafende Hoffnung den übersichtlichen, universellen Zugang zu den historischen und kulturellen Primärquellen der Welt zu ermöglichen. Zum ersten mal in der Geschichte ist es möglich, die fehlende Gegenwart darzustellen. Wir können als

Archivare nicht nur denen besser dienen, denen wir traditionell immer schon gedient haben, sondern wir haben zum ersten mal auch die Mittel, unsere Sammlungen Lehrenden und Lernenden auf allen Ebenen sowie der allgemeinen Öffentlichkeit zugänglich zu machen.

EAD und verwandte Standards haben damit begonnen, eine Informationszukunft zu gestalten, in der ernsthafte Wissenschaftler und zufällig Interessierte gleichermaßen die gesuchten kulturellen Schätze finden können. In dieser nahen Zukunft werden Informationssucher auf klar gekennzeichneten Wegen von den Katalogen zu den Suchhilfen und von den Suchhilfen zu einer Fülle von Informationen in einer Vielzahl von digitalen und traditionellen Formaten gelangen. Wir schicken uns an, nicht nur den Erkennungszugang zu unseren Sammlungen zu gewährleisten, sondern darüberhinaus den Zugriff auf wenigstens ausgewählte digitale *Faksimile* der Materialien selbst zu ermöglichen.

Zwar ist der lang ersehnte Durchbruch noch nicht vollendet, und es gibt noch viel zu tun, aber wir beginnen, uns eine ungeheuer vielversprechende Zukunft vorstellen zu können, in der nie dagewesene Formen der Zusammenarbeit zwischen Wissenschaftlern, Lehrern, Herausgebern, Archivaren und Bibliothekaren möglich werden. Über die strukturierten Datenbanken aus Katalogeinträgen, Suchhilfen und den digitalen Abbildungen von Primärquellenmaterial hinaus wird es möglich sein, private und öffentliche Informationsräume zu schaffen, die die Materialien aus einer großen Vielzahl von Perspektiven interpretieren, um einer ebenso großen Vielzahl kultureller Bedürfnisse gerechtzuwerden. Archivare werden eine wesentliche Rolle spielen, wenn es darum geht, eine digital vernetzte Informationsumgebung zu gestalten, die die intellektuelle Gemeinschaft dahingehend umzugestalten verspricht, daß nun auch gesellschaftliche Gruppen an ihr partizipieren können, die vorher nie einen Fuß in ein Archiv gesetzt oder die Schätze seiner Stahlkammer bestaunt haben.

Daniel V. Pitti
Project Director
Institute for Advanced Technology in the Humanities
Alderman Library
University of Virginia
Charlottesville, Virginia 22903