

## The Integration of Internet Resources into a Library's Special Subject Services – the Example of the History Guide of the State and University Library of Goettingen

by WILFRIED ENDERLE

### I.

The Internet has created a unique type of media for scholarly information and publication, the subject-specific web site<sup>1</sup>. Sometimes subject-specific web sites reveal still striking similarities to examples derived from the world of print, like electronic journals or text archives, but it is becoming more and more obvious that this is largely due to their authors' adherence to traditional ways of thinking rather than to a lack of knowledge about the new type of media. It is beginning to get through that we have to cope with a new type of media requiring new organisational and technical models. It is crucial to understand how insufficient it would be to try to handle these new media types in the same way as their printed predecessors. Largely there are two reasons why subject-specific web sites require new approaches to dealing with them: unlike printed books web sites do not have a fixed status and even after making them available to the public they are often changed, revised, and enlarged. Once published they do not have a fixed status but are changing regularly. They are to a large degree dynamic compared to printed resources librarians are familiar with. And secondly they are representing networked information, which does not have to have a fixed place like a library building any more because it can be accessed from all over the world via the Internet. Faced with the phenomenon that subject-specific web sites of scholarly relevance has increased considerably since about 1995, librarians have to grapple with this situation and to find answers for at least three questions:

1. Should research libraries deal with subject-specific web sites at all? If yes,
2. What is the best way for making subject-specific web sites accessible?
3. How could this concept be integrated into existing library collections and services?

I do not want to debate the basic principles and policy, so I will skip the first question, but a brief introductory remark may be allowed. Contrary to printed books Internet resources are freely and immediately accessible for everyone, at least theoretically. Thus, in principle everyone and every institution has the chance to create their own virtual library as it is often called and presumed by a lot of existing web sites. If one does not consider the creation of new institutions and organisations as the most effective way of dealing with the new media of Internet resources, two types of yet existing institutions remain which are actually organising the process of information retrieval and access for scholarly research: research libraries and specialised information centres, the main producers of subject bibliographical databases and indexing services. Making Internet resources accessible would fit very well into the function subject bibliographies actually have in the research process, i.e. to inform about the whole range of scholarly valuable and published information in certain, clearly defined subject areas. But it would also fit with the functions of a research library, which basically provides direct access to documents and to primary resources. By giving access to Internet resources this function is, partially at least, also fulfilled.

Even if it is still an open question who will be responsible for making Internet resources accessible, there can be no doubt that it has to be done and I think there are good reasons why librarians should take part in this task. Librarians have been among the first to recognise that the Internet is not a digital library giving simple and immediate access to information world wide as some of the apologists of the net believe still today. Moreover, librarians rather quickly recognised the shortcomings of the major search engines for scholarly use. And they participated in an early state of developing new and better models for giving scholars access to find really relevant Internet resources. And a last and, perhaps, decisive argument may be mentioned: for the humanities and social sciences in particular research libraries will remain indispensable for research as the main depositories of printed books. Probably - besides archives - libraries will remain the most important information providers for all scholars working in these fields. College and undergraduate teaching, in particular, may shift more to networked resources than research in the humanities. But every scholar engaged in research will depend mainly on traditional source materials which libraries and archives usually possess to a large extent. Nevertheless, current debates and research publications may shift partially to the new subject-specific websites. Having outlined these developments roughly it seems appropriate for a research library to integrate Internet resources into its existing services, since they are part of scholarly publication and communication. Thus, the main task for librarians in integrating Internet

resources is to reconcile the demands for new technical and organisational solutions subject-specific websites require with the traditional function of a research library<sup>2</sup>.

Assuming that subject-specific websites will be important even for scholars in the humanities and that libraries should therefore make them accessible, librarians need to answer two questions: how to give access to Internet resources? And how to integrate them into the main function of a research library?

There are two professional ways for making subject specific Internet resources accessible: to implement limited area search engines or to create a subject specific gateway. Both approaches can be combined and, this may, perhaps, be the solution for the future<sup>3</sup>. For now, I think, the most promising way is the concept of subject gateways. It provides the only opportunity for a scholar to get an overview of existing Internet resources of relevance for his specific research field rather quickly. At the moment general search engines may be helpful for scholarly research only in case of very specific questions. First of all, I will make this argument clear with a brief survey of current projects and nation-wide concepts (chapter 2). The concept of a subject gateway, its main criteria, will be then exemplified with the History Guide, a special subject gateway developed at the State and University Library of Goettingen (chapter 3). Then I will try to reflect the experiences we have had in building the History Guide to discuss some general problems of the subject gateway concept as well as the issue of the integration of such a new service into the services of a library that already exist (chapter 4).

I will try to make it quite clear that on the one hand we are still at the beginning of a long and difficult process with still some promising achievements yet made but on the other hand with actually much more shortcomings and troublesome problems without convincing actual solutions. I will point out some major issues as I see them and - to avoid any misunderstandings - I will deliberately do this from a relatively conservative point of view of a librarian. My main interest is to get solutions, which will strengthen in the long run the traditional function and role of a research library by adapting the new media of Internet resources in an adequate manner.

## II.

The success of the subject gateway concept can be shown by a brief survey of the most important projects. In the United Kingdom the Resource Discovery Network (RDN) has been under way since January 1999 aiming at a nationwide model with distributed subject gateways as part of so-called hubs for giving access to high quality Internet resources for specific subjects<sup>4</sup>. In most cases these hubs have been built on existing subject gateways, sometimes, as is the case in history, a new hub has been built in competition to an existing service<sup>5</sup>. Similar concepts exist in Finland and in the Netherlands, which aim at systems organised at a national level. DutchESS, the Dutch Electronic Subject Service, is much more centralised because the national library of the Netherlands, the Royal Library at The Hague, is organising it around a central database<sup>6</sup>. The Finnish Virtual Library Project is building up distributed subject gateways within an Information Gateway for selected Internet resources covering the whole range of subjects<sup>7</sup>.

In Germany research libraries responsible for special subjects within the system of the so-called Sondersammelgebiete, the special subject collections program funded by the Deutsche Forschungsgemeinschaft, the German Research Foundation<sup>8</sup>, are developing subject gateways. The organisational framework is an initiative funded by the Deutsche Forschungsgemeinschaft, aiming at the erection of virtual subject libraries centred around the special subject libraries. The foundation of the concept of these subject gateways built as part of these virtual subject libraries has been laid by a project of the State and University Library of Goettingen, the so-called Sondersammelgebiets-Fachinformationsprojekt (SSG-FI) which has resulted in the development of four subject gateways, namely for Anglo-American language and literature, earth sciences, history, and mathematics<sup>9</sup>. Other libraries are in the process of building subject gateways for psychology<sup>10</sup> and technology<sup>11</sup>; and in preparation are subject gateways for political sciences, social sciences, economics, theology, Dutch studies, art history and oriental studies.

In Australia as in the Netherlands it is the National Library, which is organising the Australian Subject Gateways<sup>12</sup>. In the United States a national concept of distributed subject gateways does not exist and I will not characterise CORC<sup>13</sup>, the OCLC project in this field, as a subject gateway but Signpost<sup>14</sup>, the subject gateway of the Scout project<sup>15</sup> located at the University of Wisconsin-Madison, has emerged as an internationally renowned one. All these projects and subject gateways existing today emphasise the current need

and lasting importance of this concept. Whether or not it will be an appropriate solution in the long run will be discussed later.

Besides, there is not only a broad range of subject gateways available, but with DESIRE funded by the European Community a project that tackles the conceptual and theoretical framework of making Internet resources accessible<sup>16</sup>. DESIRE's Information Gateways Handbook<sup>17</sup> is one of the projects making important results available to all individuals and institutions trying to build gateways of their own. Moreover, several projects are under way aiming at co-operation between different subject gateways. IMESH, the International Collaboration on Internet Subject Gateways, is an open forum for discussing conceptual topics of the collaboration of subject gateways<sup>18</sup>. And the Isaac network, a project conducted by the team of the Scout report, already offers a working technical solution for an integration of distributed subject gateways<sup>19</sup>. Renardus, like DESIRE is funded by the European Community, is still in its first stages<sup>20</sup>. It tries to develop a model for the integration of distributed subject gateways in order to create a European access point for searching scholarly relevant information on the Internet. Regarding the theoretical work as well as the subject gateways already existing, there can be no doubt about the practical impact of the concept and the main principles of a subject gateway. The latest issue of the „online information review“, edited by Traugott Koch, offers a first résumé of recent developments in addition to the DESIRE handbook already mentioned<sup>21</sup>.

But there are still unsolved problems, partly due to the imperfections of the Internet itself, namely its lack of organisation and its technical problems, partly due to the missing integration into the functions and services of a library. I do not want to discuss the issue of the actual shortcomings of the Internet itself. This would go far beyond the scope of this article, which focuses on libraries' issues. But one has to keep in mind that the inherent problems of the Internet will have an impact on the concepts and the opportunities librarians will have in dealing with the problem of integrating Internet resources into their services. It goes without saying that in a world of networked information libraries will depend on external technological developments and constraints more than ever before. Everything librarians have already achieved or are thinking about can only be part of an ongoing process of technological and organisational change and no one is able to foresee its outcome over time.

## III.

Before discussing the subject gateway approach in general I will try to sum up the state-of-the-art with its specific achievements as well as its shortcomings with the example of the History Guide located at the State and University Library of Goettingen<sup>22</sup>.

The History Guide is one of four subject gateways that have been built since 1996 as part of the so-called „Sondersammelgebiets-Fachinformationsprojekt“, a specialised subject gateways project funded by the Deutsche Forschungsgemeinschaft, which I have already mentioned<sup>23</sup>. First, a short remark about the background of the project: why did we create the History Guide and further subject gateways for specialised subject areas at Goettingen? The answer is relatively simple. Because we have a nation-wide responsibility for these subject areas. Among others the State and University Library of Goettingen is responsible for the special subject collection of Anglo-American history, politics, and language and literature. Besides its functions as a university library it acts as a kind of special subject library with nation-wide responsibilities for interlibrary loan and document delivery due to its outstanding collection in this area partially acquired with financial contributions of the Deutsche Forschungsgemeinschaft. The development of the History Guide since 1996 is a direct result of this special function as a kind of national subject library for Anglo-American history which obliges the library to acquire the printed literature, monographs and periodicals, in this field as complete as possible in order to be able to fulfil the information needs of every scholar in Germany doing research in this area. As soon as, especially in North America, subject-specific websites of value for historians came into being, the question arose how to deal with this new type of media. Without this special national responsibility and the goal to keep the collection in this field as complete as possible the necessity to build a subject gateway would not have been as urgent as it was.

The History Guide comprises about 1,450 meta data records. It offers two ways of accessing the collection: firstly via browsing a classification, and secondly via a local search engine. The most important way of access is the subject catalogue supplemented by a catalogue of source types. Second comes a local search engine giving a user the same search functionalities as a traditional online catalogue. After accessing, for example, the subject catalogue, a user will be able to navigate in a subject classification. We use the same subject classification for Internet resources and for our online catalogue, namely a locally developed classification, the Goettinger Online Classi-

*The Integration of Internet Resources into a Library's Special Subject Services – the Example of the History Guide of the State and University Library of Goettingen*

fication. We opted for the Goettinger Online Classification because an international subject classification does not exist in the discipline of history.

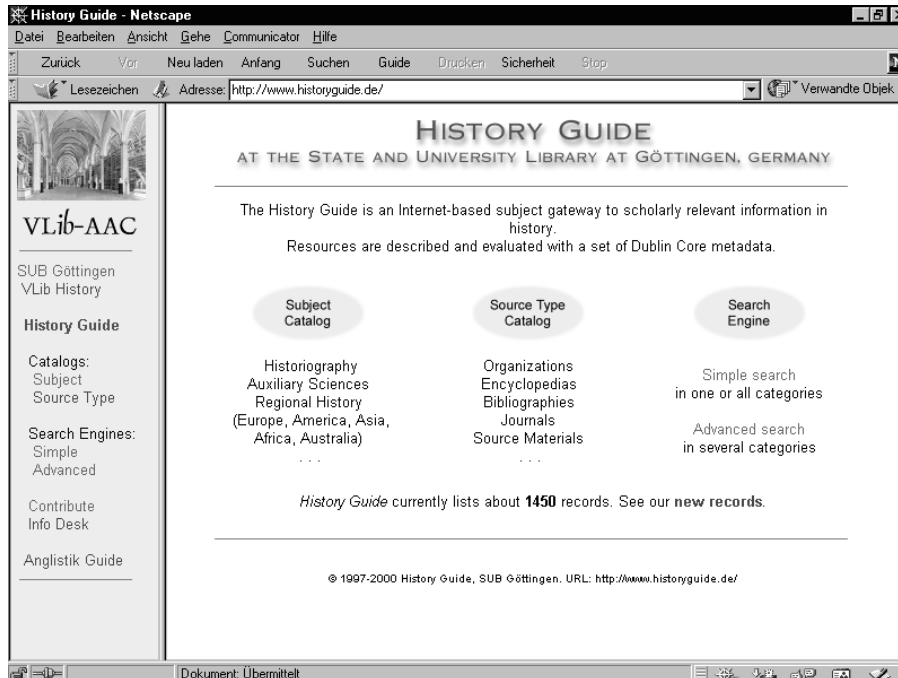


Figure 1: Homepage of the History Guide

After using one of the search facilities of the History Guide the user gets a short title list with first information about the resources. In this case I am using the subject catalogue as entrance searching for resources about the American civil war history. The short title list I will receive after browsing through the classification contains first information like the title, subject classification and the URL in order to get as fast as possible to the web site.

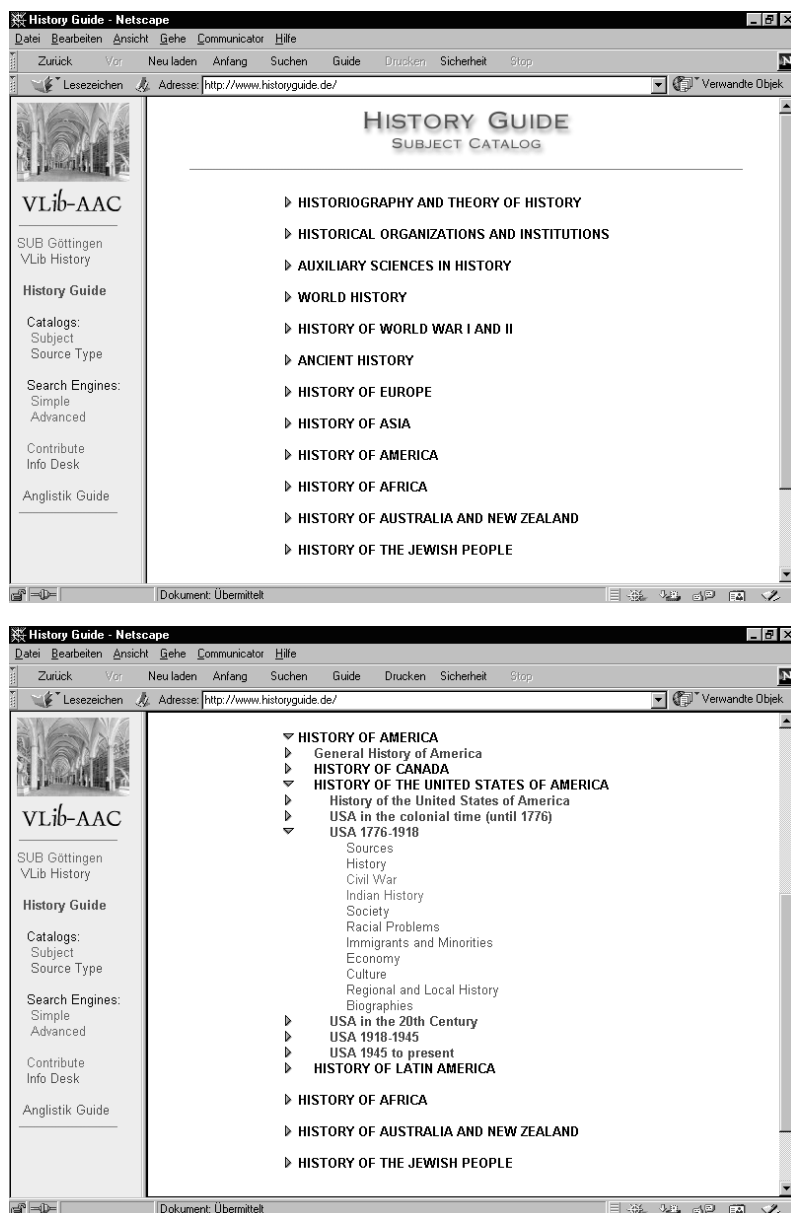


Figure 2: Subject Catalogue



*The Integration of Internet Resources into a Library's Special Subject Services – the Example of the History Guide of the State and University Library of Goettingen*

In case a user wants more information he can also retrieve the full meta data template. Probably the most practical and important feature is a description of the site, but information about authors, editors, the institution responsible for the server or statistic information like the number of back links to this specific site may also be of value. Finally, when clicking the URL and accessing the Internet resource directly, a scholar does not go on an unpredictable voyage in cyberspace with all its specific dangers like time consuming, misleading or dead links, and so on, but on a voyage prepared and instructed by a guide, built and maintained by other historians and librarians.

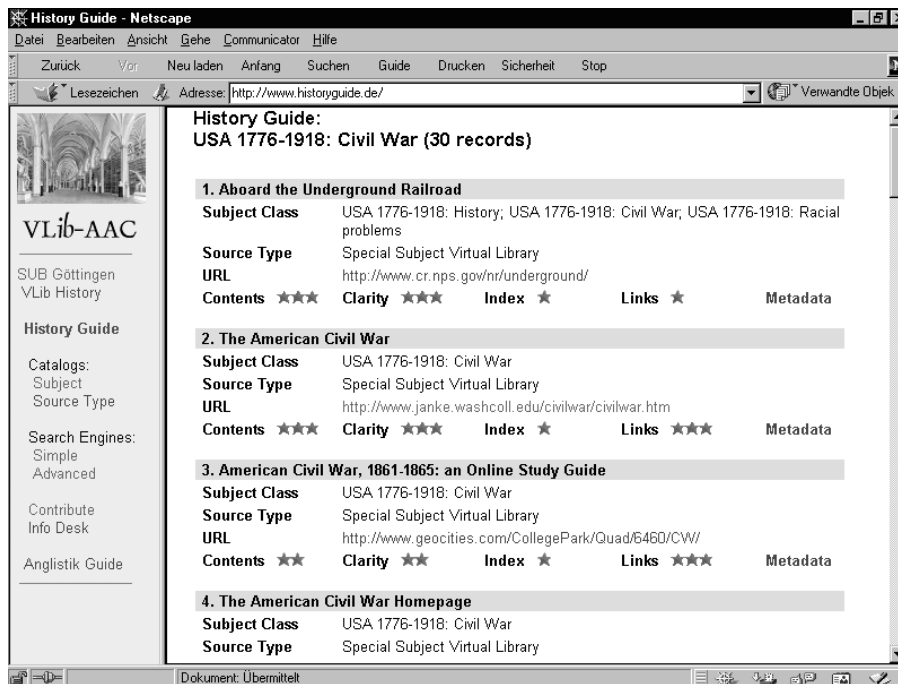


Figure 3: Short title list

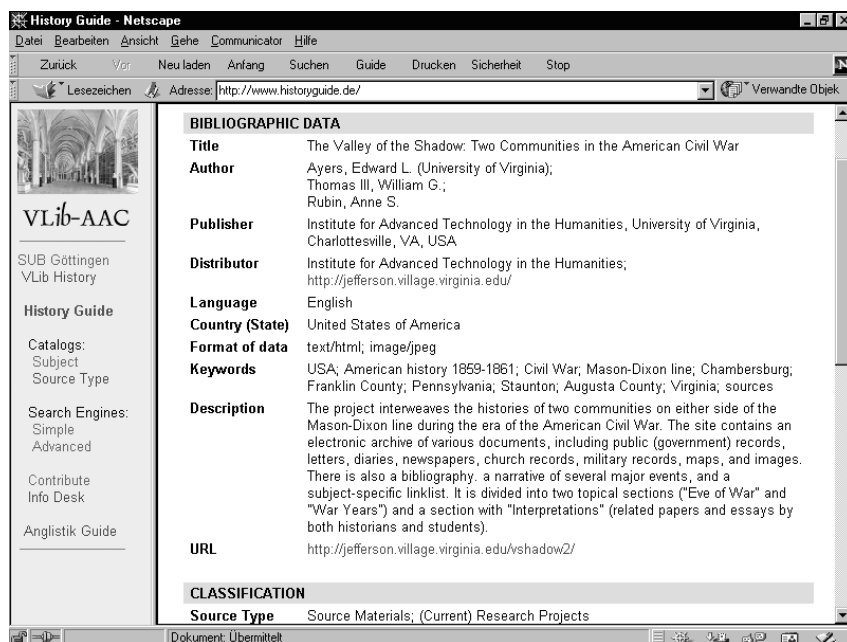


Figure 4: Meta data template

In other words: the History Guide gives access to a coherent information space not dissimilar to a good collection of books but yet without possessing any of the items. The information space made available by the History Guide is therefore independent of the library. The library cannot give its users any guarantee how this information space will develop, whether individual websites will be available in the future or how they will change their scope, size and so on.

To sum up: the History Guide in its current state can be described as a classical subject gateway because it meets the main criteria of the concept<sup>24</sup>:

1. The History Guide has a clearly defined scope, which is in accordance with the scope of the printed collections policy.
2. All resources are described by standardised meta data according to the Dublin Core.

3. All the resources described by meta data have been researched and evaluated intellectually and are of high quality. This means the resources are of value for scholarly research in one way or another. In the course of this project we have gained experiences and have developed methods for both researching and describing Internet resources.
4. For the administration of these meta data we use a database system well known in German research libraries with extensions developed by us. It is the Allegro database system developed at the University library of Braunschweig.
5. A subject catalogue gives access to the meta data via a special classification scheme. The user can browse through the classification and use the History Guide like a traditional printed bibliography.
6. A search engine is, of course, also available for searching via descriptors, authors or keywords of the title and others.

The History Guide is a new service of the library prompted by the rise of a new type of media. Thus, our library faced the problem of integrating this new service and the new type of media into its existing functions and services. But, moreover, it faced the inherent problems of the new media itself. So, the decision to build up a subject gateway gained a dynamics of its own. The History Guide was in a certain sense the nucleus of another project still under way, the development of a so-called Virtual Library of Anglo-American Culture (VLIB-AAC)<sup>25</sup>. For a research library with a main focus on the humanities this was only a small step because it is beyond any question that the printed book will remain central for any research in these disciplines. A virtual library that gives an integrated access to both types of media, to the printed monograph and periodical as well as to subject-specific websites and electronic journals seems to be a direct result of the needs of the user and the interests of the library.

This main concept of the Virtual Library of Anglo-American Culture will become clear from its home page. Access to all subject relevant resources of the library is united in a single home page, a kind of sub site of the library's main homepage. Access points are offered to monographs, periodicals, databases - mainly bibliographical databases - and Internet resources. A technical integration of some of the services is also intended.

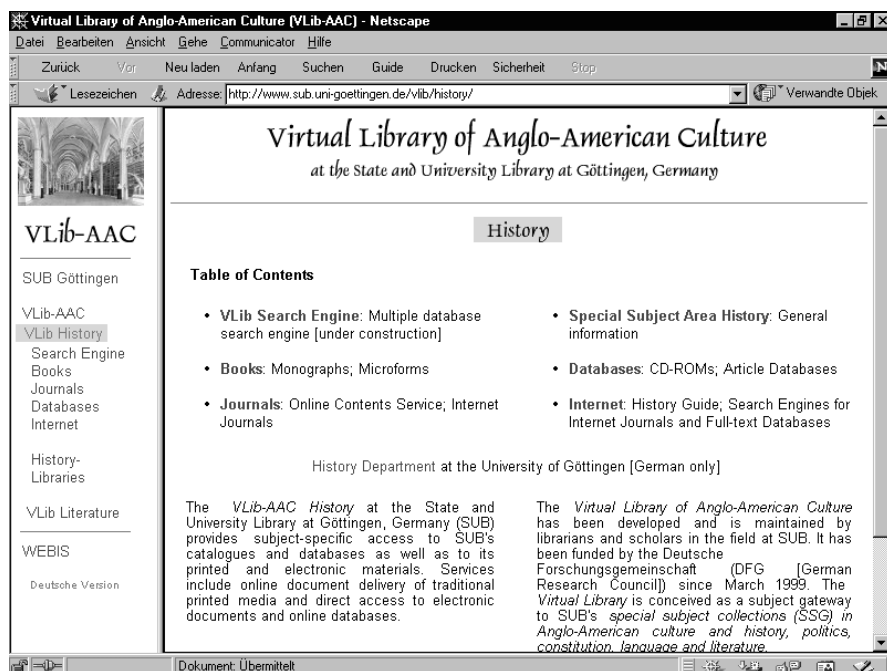


Figure 5: Virtual Library of Anglo-American Culture: History (VLIB-AAC: History)

But much more important than only a technical integration is to have a conceptual basis for subject integration. A short example what this means will be given. When a user accesses our VLIB-AAC: History home page and browses through the classification systems, if he, in other words, is doing a subject search, he will find the same system regardless whether he is looking for monographs, as a scholar traditionally does, or whether he is searching for Internet resources. What we have achieved already is an integration of contents and this is, as I will again emphasize, the basis for any technical integration of value. I will come back to this argument later.

*The Integration of Internet Resources into a Library's Special Subject Services – the Example of the History Guide of the State and University Library of Goettingen*

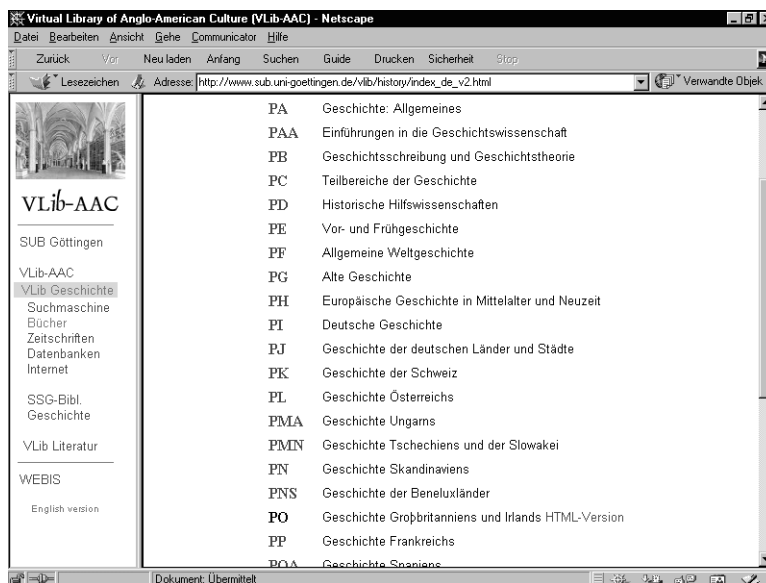
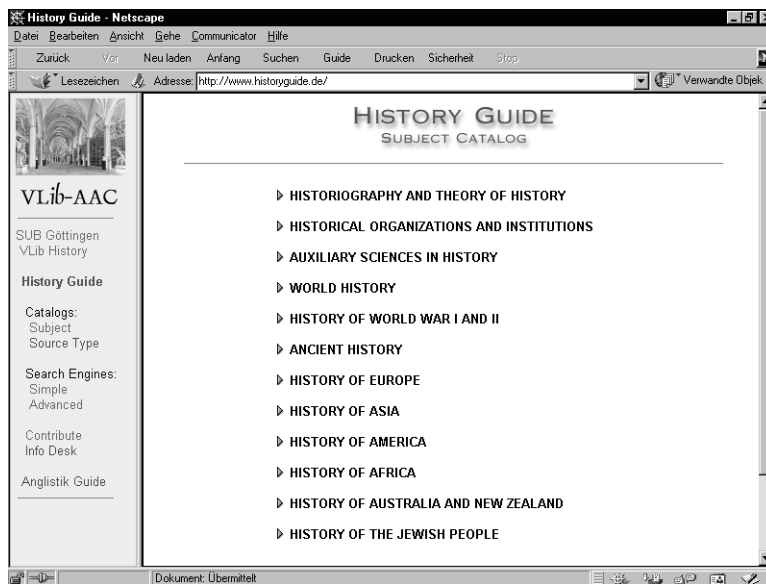


Figure 6: Subject access to monographs and to Internet resources

The History Guide initiated a development and brought up the idea of creating a subject portal for historians in which Internet resources are only one part among others and their value for actual research is considered only a minor part. I have briefly outlined this development, because I think it is a rather typical example for a general trend towards subject-based portals. I want to take the example of the History Guide and the VLIB-AAC to discuss some of the problems of the transition to a virtual library as well as general trends. Even if the project of developing and building the History Guide is successfully finished in a sense, we have meanwhile realised that we have now to approach further problems resulting from the still existing shortcomings of the subject gateway concept itself and partially from the still incomplete integration into the libraries services. So, I will try to discuss both of these issues now, the problems of integration and the still existing shortcomings of the subject gateway concept.

#### IV.

The issue of the integration of Internet resources into existing library services has often been reduced to its technical aspects, to the question how distributed databases can be integrated via standardised interfaces like Z39.50 or other protocols. This is, of course, a serious and important question, but it does not reflect how intricate the situation is in general. The technical integration is, on the contrary, a minor problem in comparison to the organisational tasks and the conceptual modifications research libraries have to make in order to fully integrate Internet resources into their daily tasks.

The topic of integration includes at least the following issues:

1. The issue of library policy: how can Internet resources and the main mission of a research library be brought together in a coherent model?
2. The conceptual issue of refining the model of subject gateways: which further developments are necessary to ensure that subject gateways will remain the adequate solution for making Internet resources accessible in the long run?
3. The internal issue of library organisation: how can new services be integrated into the existing organisational model?
4. The technical issue of integrating different databases.

There is probably no doubt that the emergence of networked resources will require academic libraries to redefine their role within the broader field of scholarly communication, information, and publication. I think it is a good approach to do this by considering the main functions of a library and how it can be transferred into the new media age. The main function of a library is to provide direct access and to provide permanent access to documents, to primary information regardless of their physical type. A subject gateway does not meet this criterion. It gives, of course, direct access to documents but it does not secure permanent access. At its present conceptual state a subject gateway does not seem to fit perfectly for a library. One can even argue whether a library really is the appropriate institution to host and maintain subject gateways. In a sense, one has to admit that other research oriented service providers who are creating indexing services are as well as research libraries free to create their own subject gateways.

But the plain fact of competing interests should not prevent research libraries from engaging themselves in this field. It is true that Internet resources are not restricted to a local collection or place. And in the past giving access to information was not only and at all a task for libraries but also for bibliographic and indexing enterprises. But, at the same time, giving direct access to a document, what libraries exactly do in creating subject gateways by themselves, is a genuine function of any library. The fundamental difference between giving access to Internet resources and giving access to books is that in the first case it can be done only via bibliographical access without any collecting of the actual items.

Exactly at this point we have to ask whether this is a sufficient way of integrating Internet resources into the functions of a library? I think not because the fundamental function of a library, the function of securing long-lasting access to documents, is excluded. So, we have to think about how subject-specific websites can be integrated within the main function of a research library? When creating subject gateways research libraries have to put emphasis on their fundamental function. Giving access to Internet resources is one thing. It is logically a necessary first step. But it does not suffice in the long run. The main task for a library should be to guarantee long-lasting access to Internet resources. As far as I see this fundamental issue of integrating Internet resources into a library's functions is not dealt with at the moment. Therefore we have not very much experience how to meet the requirements for this task.

Generally, I see two ways to achieve this purpose: first: Internet archiving<sup>26</sup> and second: the creation of consortia. Internet archiving is not just mirroring because mirroring does not meet the requirement of preserving information. Internet archiving means to preserve different versions of subject-specific websites. That this can work in principle is shown by the Internet Archive in San Francisco, initiated by Brewster Kahle<sup>27</sup>. I refer to this example not because I think it really convincing trying to archive the whole Internet. From a scholarly point of view it is not worth archiving. It is to a large extent sheer communication and simple information and only for a very little part relevant publication. But the idea itself has its merits. And the information space created by a subject gateway can be a good starting point for archiving relevant Internet resources. But Internet archiving is only one way. Another way can be to negotiate with some of the most important content providers in order to reach agreements about the future availability of their subject-specific websites and to make sure that they are archiving their own contents in a reliable manner. This will be necessary because archiving of Internet resources comprises lots of technical difficulties. One of the main problems, for example, is the invisible web, in particular the integration of information stored in database systems. To sum up: it will be necessary to create organisational infrastructures as well as new technical solutions for distributed Internet archiving in order to really integrate Internet resources into a library's mission.

If libraries can do this, then, I am convinced, the creation and maintenance of their own subject gateways will make sense. Together with other players in this field they will do a job perfectly fitting with their traditional mission. Moreover, if libraries concentrate their work on this aspect, they will really choose the path of co-operation and collaboration instead of competing with other institutions. This is important in the future because the maintenance of subject gateway is, I think, still an unsolved problem. The clear definition of their own role in the business of giving access to Internet resources may strengthen the position of research libraries in the long run.

I do not want to argue that every library should create collections of Internet resources of their own by mirroring and archiving. This would be a very traditional and insufficient method of dealing with new Internet worked resources like websites. But librarians have to create organisational models on a national or international level to secure the archiving of subject specific collections of Internet resources. And, I think, models of distributed responsibilities would fit best to the type of networked resources we are dealing with. The term collecting and collection will gain a new meaning for librarians. It would



not be necessary to overdo the organisation of distributed archiving. Collections can overlap as they do in the world of printed resources. There could be national collections for archiving purposes as well as subject oriented collections where the access for scholarly use is much more central. There could be different consortia, organised for different user groups and purposes and so forth. How this will be organised depends on local, national and subject interests. But it is crucial that such collections will be built in the future to secure the long-lasting access to materials whose future availability is very insecure today. And that the librarians do not forget what they paid for. To collect and not to dive into and get lost in cyberspace and an uncritical belief in new information technologies.

But to get there we will have also to refine the concept of subject gateways itself which will be the basis for this new type of collection building. One of the basic problems of any subject gateway is that of maintaining it because it tries to make a continuously changing world of information accessible. Considering the present state of art of subject gateways, all of them have an inherent tendency to collapse. Some sooner, some the later, but it seems to be inevitable in the end. Every subject gateway, which is continuously enlarged, requires more effort to maintain it up to a point where the amount of work for maintaining it will not allow any additions of new meta data records. It is not, as is often argued, the problem of the non-permanent Internet addresses. What makes the maintenance of subject gateways really complicated is the dynamic character of the websites. That is why the websites have to be checked regularly to keep the meta data information up-to-date. And sometimes this amounts to almost the same work as creating a new record.

To find solutions for this problem will be crucial for the future of the subject gateway concept. Besides Internet archiving the key for a solution has to be found in automated processes based on robots continuously checking the specific websites of the information space of a subject gateway. Such robots should be based on methods to measure substantial changes in files, the amount and size of files of a subject-specific web site and directory structures. Such processes should be supported by meta data standards allowing the author of a web site to include information about the site. So, by checking this meta data information, a robot can recognise more precisely if a substantial change has occurred in accordance to parameters definable by the owner of a subject gateway. Given this kind of meta-information robots can make assumptions about the frequency of revisits appropriate for a particular site<sup>28</sup>.

While this is a fundamental issue for the subject gateway concept to be established in the long run, it is also important to refine the rules and standards actually applied. And this requires to professionalise the production of meta data, of describing websites according rules and standards in accordance with existing standards and the professional level librarians are familiar with. This is also important because the creation and maintenance of subject gateways requires more collaboration with other institutions and individuals than librarians are used to. And efficient collaboration and cooperation requires well-defined rules and standards. Rules to make sure that all partners are creating compatible data, which can be easily exchanged between different institutions and subject gateways, either off-line or in a distributed database environment. This is no doubt a field where librarians are traditionally competent to work in. The standardised description of information resources and documents marked the beginning of their professionalisation at the turn of the last century and is still fundamental for the profession today. Three areas have to be distinguished for a standardised description of Internet resources: the characterisation and description of Internet resources; rules of cataloguing; and the data format<sup>29</sup>.

As far as I can see librarians have neglected the field of characterisation of web resources, although it is the basis for every standardised description of subject-specific websites. What really is a subject-specific web site? Is it identical with a web server in a technical sense? Usually not. But how can it be defined? It can have sub areas, sub home pages with an information identity of its own and so on. How should they and their interrelationships be described? What shall we do with a collection of traditional sources or monographs available in digitised form within a site? How does one deal with collections of pre-prints and working papers? A first step to reach a better understanding of these questions was undertaken by the World wide Web Consortiums' Web Characterisation Activity, which started in November 1998. The latest published result is a working draft dating from May 24, 1999 „Web Characterisation Terminology & Definitions Sheet“<sup>30</sup>. The working group tried to define the terms: web resource; web site and web page. They did it from a more technical point of view. From a librarian's point of view I think, one should add content criteria. A web site is not only defined by an Internet address and files pertaining to a consistent directory structure but also by the provided content. Therefore a web site has to be defined via its content and how the structure of this content is. It is also useful to classify the types of information. It is not easy to define websites clearly but an adequate definition is fundamental for any rules of cataloguing as well as for any collection or archiving policy. If such characterisations were available and

commonly agreed it would be also easier for authors of websites to design and layout them at the same time. A possible result might be the emergence of layout standards making it much easier for librarians to evaluate and describe the sites by their meta data.

We also have no commonly accepted cataloguing rules for subject-specific websites until now even if OCLC has done some work in this field<sup>31</sup>. To avoid any misunderstandings: I do not want to argue for a complex set of rules comparable to AACR but rather for some short definitions in accordance to the defined characterisation of websites which ought to be publicly available and agreed upon. Subject gateways have already reached a stage where international exchange of records is a common issue. Therefore the basic rules for cataloguing Internet resources should be available. This is, in my opinion, an indispensable part of any professionalisation of the subject gateway approach.

Compared with the two questions discussed already, the third issue, the issue of the data format, is in a good shape. It will therefore be sufficient to refer to the current state of the discussion about standardised meta data. Dublin Core is the standard most widely used in subject gateways and has a good chance of international recognition, due to the circumstances that the discussion about meta data arose at the same time as the first subject gateway approaches came into being<sup>32</sup>. But, it is important, not to narrow down the question of standardisation and rules to the issue of meta data. The issue of the characterisation and definition of websites and of defining internationally acknowledged cataloguing rules are equally important.

Another important feature is the organisational integration within the existing infrastructure of a library. Considering the example of the VLIB-AAC project at Goettingen, it is obvious that the organisational infrastructure of the library is challenged in a very fundamental way by the existence of the new special subject service. The example of the VLIB-AAC at the State and University Library of Goettingen shows in what ways traditional universal libraries change their character. What used to be a coherent institution is beginning to divide itself more and more in different virtual subject libraries. The technical possibility of tailoring out the services of specific subjects via a web interface results in dividing a universal library into different units with special purposes and services. For the library itself this will lead to modifications of the internal organisation in the long run. It will become increasingly important to have subject teams organising the virtual libraries in close connection with other libraries, research institutions and professional organisations. In such an

environment it will become more obvious that research libraries are not autonomous and independent but part of the broader infrastructure of an academic discipline.

At a more basic level it will be necessary to integrate the work of making Internet resources accessible into the practical workflow of library. The experience of our project has shown that it is relatively difficult to come to a strictly modularised workflow. The best way is to have a team collaborating very closely. The reason is that a large amount of work in producing meta data records is in the evaluation of a web site. Obviously, it is most efficient to let the same person who is evaluating a site produce most of the meta data. The difficulty in modularising the workflow is largely due to missing layout and description standards of subject-specific websites. Thus it is often a laborious procedure to get all the necessary information about a site such as the author, the commission or institution actually responsible for publishing it, or the size of the site and so on.

Apart from the problem of integration into the workflow we must also consider a broad integration of subject indexing and classification. In other words: the integration of a subject gateway into other resources and services of a library makes it necessary to approach the issue of subject classification of different resources. Usually, a library offers their users a subject catalogue in addition to existing indexing services and bibliographical databases. With the emergence of the World Wide Web and its browsing facilities the access via a classification scheme has gained much more relevance, at least in the humanities where such an approach has a long tradition derived from the printed bibliographies. Having put together different services in a subject portal on the web, such as library catalogues, indexing databases, subject bibliographical databases, and a subject gateway as we have done it in the VLIB-AAC project, one recognises how widely the subject classification of the different services differ. Thus it is quit clear that the integration of subject classification from different systems will be a major task of the future. For any integration with other services, it is important to consider as early as possible the topic of subject classification of Internet resources. This will differ from discipline to discipline but it will be generally helpful to apply classification schemes commonly used in a discipline or by the institution hosting the subject gateway. From the user's point of view a real integration will only be achievable this way. Our first step was to apply the same classification scheme for Internet resources we are using for our monographs. To create concordances with other systems is far beyond our current reach but should be logically the next step.

The technical integration of different services will be only useful if concordances between different subject schemes exist. Otherwise the quality of searches done with a meta search engine integrating different services like indexes, library catalogues and subject gateways will be at a very low level. Usually, a patron has only the possibility to use free key words from the records. The quality of the results of such a search will not differ greatly from that of an old generation Internet search engine. The user can never be sure whether he has really retrieved all relevant information available in the integrated databases. In the end he will have to approach the services one by one to get a high quality search. This is why I think that the technical integration is a minor problem. I do not underestimate the complexity of the technical issue but to tackle the issue of the technical integration requires solutions for subject classification as well as clear standards of cataloguing first.

So I want to argue that we should not care too much about the issue of the technical integration. The technical matters of integration are in principle solved. Currently there are three protocols which can be used: LDAP, HTTP and Z39.50. All of these protocols are already implemented in different projects. So, experiences are available. The main problem of a virtual library is that these protocols were designed for different database applications. Z39.50 for traditional online catalogues, HTTP for web applications in general and LDAP was developed in the context of a WHOIS application. Today it is possible to integrate different subject gateways as the Isaac network shows. It will be of course necessary for a sophisticated integration of a subject gateway with a library catalogue or a subject bibliographical database to make detailed agreements about the exchange format of the meta data, about the indexing rules of the different databases and the technical interfaces. But these are only technical problems and there are now major obstacles for their implementation with the exception of, may be, missing financial resources and available staff.

Finally, I will again emphasise that the main issue in integrating Internet resources into the services of a library is not the technical one but the question of an amalgamation with the main mission of a research library. I think that we have with the subject gateway concept an approach perfectly matching the interests of scholars as well as of librarians. But to ensure that the subject gateway concept can be established in the long run librarians have to tackle three problematic tasks at least:

1. The organisational and technical infrastructure for guaranteeing long-lasting access to Internet resources has to be developed. In the world of networked information new models of collection building and archiving have to come into being in order to reconcile the traditional mission of a research library with the new media type of the subject-specific web site.
2. There is an urgent need to create technical and organisational solutions for the problem how meta data referring to dynamic resources can be regularly updated. Without such solutions the long-term maintenance of subject gateway cannot be secured.
3. Librarians have to define an internationally acknowledged set of rules and standards for describing subject-specific websites which should include definitions for characterising websites and rules for cataloguing them in addition to a standardised meta data format already available with the Dublin Core.

So, my main argument as a librarian is that we have to keep in mind what a research library basically has to do, that is to secure direct and permanent access to primary information items. This is in my view the real challenge librarians are faced with in integrating Internet resources into the functions and services of a research library.

#### REFERENCES

- 1 See in more details: Clifford Lynch: On the Threshold of Discontinuity: The New Genre of Scholarly Communication and the Role of the Research Library. Paper delivered at the ACRL Ninth National Conference April 8-11, 1999, Detroit, Michigan.
- 2 Clifford Lynch: Searching the Internet. In: Scientific American, March 1997 <<http://www.sciam.com/0397issue/0397lynch.html>>.
- 3 Some interesting notifications about new trends in web searching gives Chris Sherman: The Future Revisited: What's New with Web Search. In: Online, May 2000 <<http://www.onlineinc.com/onlinemag/OL2000/sherman5.html>>. See also: Soumen Chakrabarti / Martin van den Berg / Byron Dom: Focused Crawling: A New Approach to Topic-Specific Web Resource Discovery, 29 March 1999 <<http://www8.org/w8-papers/5a-search-query/crawling/index.html>>.

*The Integration of Internet Resources into a Library's Special Subject Services – the Example of the History Guide of the State and University Library of Goettingen*

- An example for a limited area search engine is ARGOS - Limited Area Search of the Ancient and Medieval Internet <<http://argos.evansville.edu/index.htm>>. Background information gives: Anthony Beavers: Evaluating Search Engine Models for Scholarly Purposes: A Report from the Internet Applications Laboratory, 1999 <<http://www.icaap.org/TheCraft/content/1999/beavers/index.html>>.
- 4 <<http://www.rdn.ac.uk>>
- 5 See History online resources of the Institute of Historical Research <<http://ihr.sas.ac.uk>>.
- 6 Marianne Peereboom: DutchESS: Dutch Electronic Subject Service – a Dutch national collaborative effort. In: Online Information Review, Vol. 24 (2000) p. 46-48. – See also: <<http://www.konbib.nl/dutchess>>.
- 7 <<http://www.jyu.fi/library/virtuaalikirjasto/engvirli.htm>>.
- 8 A good access point for the services of the special subject collections program is WEBIS, the web information system of the special subject collections <<http://webis.sub.uni-hamburg.de>>.
- 9 <<http://www.sub.uni-goettingen.de/ssgfi/index.html>>. – See also the published documentation of the project: Das Sondersammelgebiets-Fachinformationsprojekt (SSG-FI) der Niedersächsischen Staats- und Universitätsbibliothek Göttingen. GeoGuide, MathGuide, Anglo-American History Guide und Anglo-American Literature Guide (<<http://www.SUB.Uni-Goettingen.de/ssgfi/>>) Dokumentation, Teil 1 (=dbi-materialien, Bd. 185) Berlin 1999.
- 10 <<http://fips.sulb.uni-saarland.de/fips.htm>>.
- 11 <<http://www.tib.uni-hannover.de/vifatec/>>.
- 12 <<http://www.nla.gov.au/initiatives/sg/>>. See also: Roxanne Missingham: Portals Down Under: Discovery in the Digital Age. In: Econtent, April/May 2000, p. 41ff.
- 13 <<http://www.oclc.org/oclc/research/projects/corc/>>.
- 14 <<http://www.signpost.org/>>.
- 15 <<http://scout.cs.wisc.edu/>>.

- 16 <<http://www.desire.org>>.
- 17 DESIRE Information Gateways Handbook, Nov. 10, 1999 <<http://www.desire.org/handbook/print4.html>>.
- 18 <<http://www.imesh.org>>.
- 19 <<http://wwwscout.cs.wisc.edu/research/>>. See also: Susan Calcari: The Internet Scout Project. In: Library HI TECH, Vol. 15 (3-4) 1997, p. 11ff.
- 20 <<http://www.renardus.org>>.
- 21 Traugott Koch (ed.): Subject Gateways (=Online Information Review, Vol. 24, 2000).
- 22 <[www.historyguide.de](http://www.historyguide.de)>.
- 23 See footnote 9.
- 24 Traugott Koch: Quality-controlled subject gateways: definitions, typologies; empirical overview. In: Online Information Review, Vol. 24 (1) 2000, p. 24ff.
- 25 <<http://www.sub.uni-goettingen.de/vlib/index.html>>.
- 26 One has to distinguish the more specific topic of Internet archiving from the topic of preservation of digital materials which has a much broader scope. The later issue is approached by librarians for years. See as a case in point: Preserving Digital Information. Report of the Task Force on Archiving of Digital Information, commissioned by The Commission on Preservation and Access and the Research Libraries Group, May 1, 1996. Concerning the current discussion see Gail M. Hodge: Best Practices for Digital Archiving. An Information Life Cycle Approach. In: D-Lib Magazine, Vol. 6(1), January 2000 <<http://www.dlib.org/dlib/january00/01hodge.html>>.
- 27 Brewster Kahle: Preserving the Internet, 1997 <<http://www.sciam.com/0397issue/0397kahle.htm>>. – See also: The Internet Archive: Building an Internet Library <<http://www.archive.org/>>.
- 28 In principle such programs exist, but they are usually designed for personal use. See Greg R. Notess: Internet Current Awareness. In: Online, March/April 1999, p. 75ff.



*The Integration of Internet Resources into a Library's Special Subject Services – the Example of the History Guide of the State and University Library of Goettingen*

- 29 See also: Erik Jul: Cataloging Internet Resources: An Assessment and Prospectus. In: The Serials Librarian, Vol. 34 (1-2) 1998, p. 91ff.
- 30 Web Characterisation Terminology & Definitions Sheet. W3C Working Draft 24-May-1999 <<http://www.w3.org/1999/05/WCA-terms>>.
- 31 See: Cataloging Internet Resources. A Manual and Practical Guide, second edition, Dublin/Ohio 1997. - See also as appropriate discussion forum for this topic: Journal of Internet Cataloging, 1996- <<http://www.haworthpressinc.com:8081/jic/>>.
- 32 See: Dublin Core Metadata Initiative <<http://purl.org/dc/>>.

Dr. Wilfried Enderle  
Niedersächsische Staats- und Universitätsbibliothek  
37070 Göttingen, Germany  
Tel: +49 551 395200  
[enderle@sub.uni-goettingen.de](mailto:enderle@sub.uni-goettingen.de)