

Angelika Menne-Haritz (Hrsg.)

Online-Findbücher, Suchmaschinen und Portale

**Beiträge des 6. Archivwissenschaftlichen Kolloquiums
der Archivschule Marburg**

Die Deutsche Bibliothek - CIP-Einheitsaufnahme
Ein Titeldatensatz für diese Publikation ist bei Der
Deutschen Bibliothek erhältlich

ISBN 3-923833-68-7

Gedruckt auf alterungsbeständigem Papier ∞

© Archivschule Marburg 2002

Bismarckstraße 32, D 35037 Marburg

Konzeption der Umschlaggestaltung: Designbüro Glas, Seeheim-Jugenheim

Druck: Völker & Ritter, Schwanallee 27-31, 35037 Marburg

Inhalt:

Vorwort	7
Angelika Menne-Haritz: Internet und Archive - Die Wiederentdeckung der Strukturen	9
Anne Van Camp: Integrating Access to Archives – The RLG Experience	19
Kris Kiesling: The American Archival Community - Why We Need EAD	27
Catherine Dhérent: French experiences with the adaptation of EAD	37
Bernhard Grau: Das deutsch-amerikanische Projekt "Gemeinsames Fachkonzept Online-Erschließung" - Zur Übertragbarkeit der Encoded Archival Description (EAD) auf die archivische Praxis in Deutschland ..	49
Gerald Maier und Thomas Fricke: Digitalisierung von Archivgut als Teil eines archivischen Online-Angebots - Workflow und Werkzeuge zur Digitalisierung von Archivgut. Ein Werkstattbericht.....	67
Edgar Büttner: Das Archivverwaltungssystem BASYS als Voraussetzung für die Internetpräsenz des Bundesarchivs	93
Mechthild Black-Veldtrup: Das DFG-Projekt „Entwicklung von Werkzeugen zur Retrokonversion archivischer Findmittel“	115
Manfred von Boetticher: Kartographische Darstellung historischer Zusammenhänge im archivischen Internetauftritt - Neue Formen einer benutzungsorientierten Suchstrategie.....	129
Frank M. Bischoff: Das Projekt V.E.R.A. in Nordrhein-Westfalen - Nutzung der Internettechnologie für die Erschließung und archivübergreifende Verwaltung der Bestände	135
Bernward Helfer: Das Hessische Erschließungssystem HADIS 2000..	153
Sigrun Eckelmann: Die Archivförderung der DFG – Schwerpunkte und Perspektiven	173
Paul Bantzer: Datenbanken für Internetangebote	183
Gerhard Knorz: Visualisierung von Zusammenhängen - Von der Wissenskarte zur interaktiven graphischen Topic Map	195
Manfred Sommer: Gemeinsame Homepages und Portale - Möglichkeiten und Voraussetzungen.	219
Index:	231
Die Autorinnen und Autoren	235

Vorwort

Im Juni 2001 fand im Software-Center Marburg das 6. Archivwissenschaftliche Kolloquium der Archivschule Marburg unter dem Titel Online-Findbücher, Suchmaschinen und Portale statt. Es führte Erfahrungsberichte aus verschiedenen Archivanwendungen mit Ausblicken auf die technologischen Möglichkeiten der näheren Zukunft zusammen. Im vorliegenden Band sind die Beiträge dieser Tagung zusammengestellt.

Die Internationalisierung der Technik erfordert eine Intensivierung des internationalen fachlichen Austausches. Das Kolloquium gab deshalb in drei Beiträgen den Ausblick auf den Schwerpunkt der aktuellen internationalen Diskussion um die Nutzung des Internet für die Verbesserung des Zugangs zu Archiven, bei denen die Encoded Archival Description, eine von der Society of American Archivists entwickelte Norm, im Zentrum steht. Die große Verbreitung von EAD - immerhin sind allein über die Suchmaschine der RLG in San Francisco bereits über 20000 strukturierte Findbücher kombiniert erreichbar - markiert das im internationalen Rahmen erneut gestiegene Interesse an Findmitteln zur strukturierten Präsentation von Archivgut. Sie geht einher mit der Nutzung von XML zur Beschreibung der Strukturen. Diese Technik wird auch in den neuen Ansätzen, die in den Beiträgen aus mehreren deutschen Archiven beschrieben wurden, genutzt. Deshalb wurden mit den abschließenden Beiträgen des Kolloquiums ein Ausblick auf weitere technische Möglichkeiten eröffnet. Dabei ging es um die Nutzung von Datenbanken für die Erfassung der Angaben für Internetangebote und für ihren Präsentation und die Nutzung, der Einsatz von Portalen wurde vorgestellt und eine neue Nutzung der XML-Markierungen zur Darstellung von Bezügen mit Hilfe von thematischen Karten, den sogenannten Topic Maps wurde erläutert.

Die Beiträge des Kolloquiums werden hier in der Sprache, in der sie gehalten wurden, publiziert. Die Folienpräsentationen der Tagung sind bis auf weiteres auf der Website der Archivschule einzusehen. Zitierweise von Literaturangaben und Referenzen sind bewusst nicht vereinheitlicht worden. Ihre jeweilige Form entspricht den jeweiligen nationalen oder disziplinären Regeln und den Wünschen der Autoren.

Angelika Menne-Haritz: Internet und Archive - Die Wiederentdeckung der Strukturen

Die Nutzung des Internet in den Archiven hat in der letzten Zeit einen großen Aufschwung genommen. Die Skepsis, ob man auf diesem Weg wirklich die Öffentlichkeit erreicht, ist der Erfahrung gewichen, dass mit den neuen Instrumenten ein direkterer Kontakt mit potentiellen Benutzern möglich ist. Es wird allmählich kompliziert, den Überblick über die Internetangebote der Archive zu behalten.¹ Die Website gehört inzwischen einfach dazu. Je mehr nun also Erfahrungen mit dem neuen Medium gesammelt werden, umso größer wird auch der Bedarf nach Erfahrungsaustausch.

Das Internet ist nicht nur das große neue multimediale Lexikon, für das es zu Anfang oft gehalten wurde. Man schlägt nicht einen Begriff nach und findet dann die authentische Erläuterung, was er bedeutet. Die Nutzung des Internet geschieht auf andere Weise, auch wenn der Einstieg oft verblüffende Ähnlichkeit mit dem Nachschlagen in einer Wortliste hat. Oft wird die Eingabe von einem oder mehreren Suchbegriffen in einer Suchmaschine als Einstieg genutzt. Doch die Ergebnisse, die sie liefert, sind Adressen von Seiten, auf denen der Begriff verwendet wurde. Beim Anklicken dieser Seiten wird in der Regel keine Definition des Begriffs gefunden, sondern er erscheint in einem Verwendungszusammenhang, der ohne Berücksichtigung irgendwelcher Definitionen entstanden ist. Das ist ein großer, ein prinzipieller Unterschied zum Lexikon. Während das Lexikon dem Leser eine Mitteilung über die Bedeutung eines Wortes machen möchte, ist die Absicht der Autoren von Websites normalerweise etwas anderes. Ähnlich wie ein Zeitschriftenartikel, ein Wer-

¹ Auf der Website der Archivschule Marburg sind inzwischen über 300 deutsche Archive mit ihren Internetadressen aufgelistet.

beprospekt oder ein wissenschaftlicher Forschungsbericht sollen die potenziellen Leser aufgefordert werden, sich mit einer ihnen interessant erscheinenden Sache auseinander zu setzen. Darüber hinaus können die Leser zu einer direkten Aktion aufgerufen werden, zum Anklicken von Links, zur Rückmeldung mit EMail, zu einer Abstimmung oder zu einer Bestellung und einem Kauf.

Den Mitteilungen des Internet geht es nicht nur um Worte, sondern um Taten. Viel direkter als andere schriftliche Mitteilungen es können, wollen sie aktivieren und bieten die Möglichkeit zur Interaktion an. Das ist der Unterschied zu gedruckten, analogen Medien, die keine direkte Reaktion provozieren, steuern und entgegen nehmen können. Das Internet beruht in seinem Gehalt und seinen Aussagen sehr viel stärker als bisherige schriftliche Mitteilungen auf den Intentionen und Absichten seiner Autoren, dass eine gewünschte Aktion geschehe.

Diese Wirkung des Mediums schafft eine neue Erfahrung der Kommunikation. Da es nicht nur um die Verteilung von Informationen an ein anonymes Publikum geht, sondern weil Reaktionen möglich und gewünscht sind, tritt die Seite der Adressaten stärker in den Vordergrund. Es wird deutlich, dass die Vermittlung von Informationen in einer Kommunikationssituation erst dann erfolgreich ist, wenn die Empfängerseite sie wahrnimmt, versteht und akzeptiert, sie womöglich braucht und gerade gesucht hat, sie nutzen und auch direkt darauf reagieren kann.

Zu einer sinnvollen Arbeit mit Informationen aus dem Internet gehört es deshalb vermehrt, die dahinter liegenden Absichten mit zu bedenken, um das, was mitgeteilt wird, zu verstehen. Die Entdeckung der Absichten geschieht nicht, um sie zu enttarnen, sondern einfach um zu verstehen, was gemeint ist und was bewirkt werden soll. Danach kann man entscheiden, wie man selbst die Informationen interpretieren und aufnehmen will. Um die Glaubwürdigkeit und Seriosität von Angaben zu beur-

teilen, sucht man nach der Herkunft der Website. So ist etwa die Angabe über die Bestände von Archiven, die, wie es zunächst geschah, von Genealogenvereinigungen ins Netz gestellt wurden, weniger authentisch für potentielle Benutzer, als die Angaben des Archivs selbst, weil sie nur für die Zwecke einer bestimmten Nutzergruppe gemeint waren. Das war durchaus ein wichtiges Motiv für Archive, selbst über die eigenen Bestände zu informieren. Doch die Nutzung des Internet für Archivzugänge hat noch einen weiteren Grund. Das ist die interaktive Nutzungsmöglichkeit, das Navigieren durch die Bestände. Hier scheint das Internet wie für die Archive erfunden.

Schon mit analogen Findmitteln müssen Benutzer arbeiten. Es genügt nicht, sie nur zu lesen. Was in den Findmitteln zu sehen ist, muss interpretiert werden. Interessant, ist was dahinter steckt. Titel von einzelnen Einheiten und Gliederungsstrukturen können nur zusammen verstanden werden, weil die Überschriften der Gliederungsgruppen den Text der Titel weiter erläutern und ergänzen. Die in der Gliederung der Findbücher oder der Tektonik der Beständeübersichten abgebildeten und in den Titeln der Verzeichnungseinheiten formulierten Entstehungszusammenhänge bilden das Netz von Aktivitäten und Reaktionen, von Fragen und Antworten ab, aus dem das Archivgut als Nebenprodukt stammt und in dem man sich schlussfolgernd vorwärts bewegt. Die Struktur von Archivfindmitteln repräsentiert die dem Archivgut zugrunde liegenden Handlungszusammenhänge und stellt damit durchaus eine andere Ebene als etwa inhaltliche Sachzusammenhänge dar, die in einer Sachsystematik erfasst sein können. Diese Sicht der Strukturen als Resultate von auf einander reagierenden Handlungen gilt nicht nur für Findbücher. In den Findbüchern präsentiert die Gliederung die Zusammenhänge des Archivguts. Deshalb ist das schlussfolgernde Interpretieren auch die Methode, mit der Archivgut als Quellen interpretiert wird und für neue, unbekannte Informationen ausgewertet werden kann. Wenn man sich die Mühe

macht, Findbücher als Abbildung von Netzwerken zu sehen, dann erfährt man bei der Arbeit mit ihnen etwas völlig Neues, was noch niemand wusste und deshalb auch nicht aufgeschrieben oder jemandem anderem mitgeteilt hat.

Beständeübersichten und Findbücher, wie in der bisherigen analogen Form so auch in der digitalen, elektronischen Form sind keine Mitteilungen an Leser wie andere gedruckte Publikationen einschließlich der Nachschlagewerke und Lexika. Findmittel zum Archivgut sind Hilfsmittel für einen vermittelten Zweck oder, wie Georg Friedrich Raumer zu Beginn des 19. Jh es formulierte, so etwas wie die Landkarten zu den unbekannten Landschaften in den Beständen im Magazin des Archivs. Wie Landkarten zeigen sie die Wege, auf denen man sich seinem noch nicht bekannten Ziel nähern kann. Landkarten zeigen Ebenen und Berge, lassen erkennen, wo Flüsse fließen und welche Brücken man überqueren kann. Um zum Ziel zu kommen, kann man interessante Umwege nehmen oder der direkten Straße folgen. Man kann sich von Details am Rande ablenken lassen oder ohne rechts und links zu blicken den geraden Weg einschlagen. Genauso nutzt man Findbücher. Auch sie können mehr oder weniger zielgerichtet durchschritten werden. Details am Rande können das Interesse ablenken. Doch ist es eigentlich immer klar, was tatsächlich weiterhilft und was zusätzliche Verstärkung oder Illustration der eigentlichen Ergebnisse ist. Vor allem ist klar, was nicht weiterhilft. Das ist eine zentrale Funktion der Findmittel. Wichtiger als der positive Nachweis einschlägiger Quellen ist die sichere Erkennbarkeit von nicht erforderlichen Nachforschungen, die in Sackgassen führen würden und viel Zeit und Arbeitskraft absorbieren können. Die eindeutige Entscheidung über die aus Sicht der gerade verfolgten Fragestellung nicht erforderlichen Wege sichert die Vollständigkeit der Quellenbasis für eine Nachforschung ab. Denn erst, wenn es eindeutig ist, dass es nichts gibt, was gegen das gefundene Ergebnis spricht, kann man sich darauf ver-

lassen und es als Basis für weitere Forschungen nutzen. Solche Einschätzung erlauben nur strukturierte Findbücher, nicht aber Listen möglicher Inhalte, sei es als Register, Indizes, Deskriptoren oder in anderer Form.

Archivgut entstand nicht für Archivnutzer. Es stammt aus Kommunikationszusammenhängen, die den Benutzern im Archiv zunächst fremd sind und in die sie sich erst einfinden müssen, bevor sie die Spuren, die vor ihnen ausgebreitet werden, verstehen und für ihre Fragestellung interpretieren können. Dabei betreten sie Neuland, in dem sie sich oft vor der Fülle an neuen Eindrücken zunächst um Orientierung bemühen, bevor sie sich mit der Klärung ihrer Fragen befassen können.² Die Orientierung kann man sich nur selbst erarbeiten. Nur die eigene Fragestellung bringt Ordnung in das Gelände der Angebote. So wie Autofahrer eine Landkarte nutzen, um die Straßenverbindungen zu ihrem Ziel zu finden, Radfahrer oder Wanderer aber vielleicht gerade, um die Autostraßen zu vermeiden, ist es auch bei Ermittlung von Archivgut als Quellen die eigene Perspektive, die Angaben im Findbuch nützlich macht. Die Notwendigkeit, sich eigenständig zurecht zu finden, kann abschreckend wirken, wenn nicht erkennbar ist, was man tun kann. Sie erscheint dann als Schwellenangst, als Ausdruck der besonderen Schwierigkeiten bei der Nutzung von Archiven. Doch mit komfortabel zu nutzenden Hilfsmitteln, mit guten, übersichtlichen Landkarten, findet man selbst den Weg zu überraschenden neuen Erkenntnissen.

Die Orientierung in den angebotenen Beständen einerseits und die Auswahl der als einschlägig heranzuziehenden Unterlagen sind die zwei zusammengehörenden Seiten der Ermittlung von Quellen, die der Auswer-

² Wendy M. Duff und Catherine A. Johnson, *Accidentally found on Purpose. Information seeking behaviour of historians in archives*, Vortrag bei der Jahrestagung der Society of American Archivists, Washington 2001.

tung des Archivguts vorangehen. Eben so wenig wie die Interpretation der Quellen kann deren Ermittlung und Auswahl anderen Personen überlassen werden, wenn das Ergebnis der Forschungen tragfähig werden soll. Archivare kennen zwar das Material besser als ihre Benutzer. Doch kennen Benutzer ihre eigenen Fragestellungen besser und können deshalb selbst am besten entscheiden, was sie benötigen. Diese Entscheidung kann allerdings unterstützt werden, wenn Archivare möglichst viel ihrer bei der Übernahme und Erschließung erworbenen Kenntnisse direkt und selbsterklärend bereitstellen. Strukturierte Beständeübersichten und Findbücher erlauben den Benutzern die Orientierung in den Beständen und unterstützen ihre eigene Entscheidung, welches Material sie ansehen wollen.

Zusätzlich zu diesen traditionellen, bereits analog vorhandenen Hilfsmitteln können mit Hilfe der Internet-Technologien weitere, neue Instrumente konzipiert und bereit gestellt werden, die in der analogen Welt nicht oder schwieriger zu handhaben wären. Dazu gehören die von MIDOSAonline genutzten Zusatzinformationen, mit denen Informationen über die Bearbeitung einzelner Verzeichnungseinheiten, über ihr Schicksal vor der Übernahme oder über Bewertungsentscheidungen zum Aufrufen bei Bedarf angeboten werden können³. Dort können viele bei der Übernahme und der Bearbeitung angefallenen Informationen als Ergänzung zur eigentlichen Erschließung sinnvoll untergebracht werden. Ebenso kann man dazu die Digitalen Praktika rechnen, mit denen Nutzern die Möglichkeit gegeben werden kann, an digitalen Reproduktionen auszuprobieren, welche Verfahren für die Auswertung erforderlich sind und

³ Angelika Menne-Haritz, Das Online-Findbuch - Archivische Erschließung mit Internettechnologie, in: Angelika Menne-Haritz, Hrsg., Archivische Erschließung - Methodische Aspekte einer Fachkompetenz, Veröffentlichungen der Archivschule Marburg Nr. 30, Marburg 1999, S. 109-122, bes. S. 117-118.

welche Vorkenntnisse sie in den Lesesaal mitbringen sollten. Auf diesem Gebiet gibt es noch viele unausgeschöpfte Spielräume für archivfachliche Phantasie.

Findbücher zeigen, wo es bei der Arbeit, die zur Entstehung des Archivgutes führte, besondere Probleme in einer Behörde gab, wo sie routinisiertes Tagesgeschäft erledigte, wo der ganze Aufgabenbestand sich wandelte und wie man mit solchen Änderungen fertig wurde. Ebenso wie die Autoren von Webseiten normalerweise ihre Texte nicht nach dem Bedarf der Suchmaschinen erstellen, genauso wenig hat die Behörde Informationen als Mitteilung für potentielle Archivbenutzer aufgeschrieben. Der Klick auf die Website zeigt ihren Autor in der Aktion, lässt seine Aufforderung zur Aktion und seine eigene Intention erkennen, und genauso erlaubt das Findbuch, der jeweiligen Dienststelle über die Schulter zu blicken und sie in der Aktion zu sehen.

Beständeübersichten, die hohen Aufwand bei der Erstellung erfordern, durch ständige Neuzugänge eventuell bei ihrem Druck aber schon veraltet waren, können im Internet ständig aktualisiert und verbessert werden. Internet-Beständeübersichten zeigen zudem mehr als analoge Exemplare. Die Beschreibung der einzelnen Bestände selbst steht nicht mehr so eindeutig im Vordergrund. Mit der Nutzung des Internet ist die Tektonik der Archive als Gliederung der Beständeübersichten und damit als Gesamtübersicht über alle, mit einander verknüpften Bestände eines Archivs wieder wichtiger geworden. Der aus der Architektur stammende Begriff Tektonik bezeichnet die Struktur des Hauses, die seine Teile zu einem Ganzen zusammenhält. Die Tektonik der Bestände, genutzt als Gliederung der Beständeübersicht, hält auf die gleiche Weise das Gebäude des Gesamtbestandes eines Archivs zusammen.

Die Eigenart der archivischen Gliederungsprinzipien, nämlich die Teile eines gegliederten Ganzen in ihren Beziehungen zueinander zu zeigen, geht

auf die Entstehung der Bestände für einen anderen Zweck als die Benutzung im Archiv, nämlich für die Koordinierung gemeinschaftlicher Aktion, zurück. Der Blick über die Schulter zeigt, wie eine Aktion auf eine andere reagiert und wie dabei schriftliche Aufzeichnungen eingesetzt wurden. Er zeigt Absichten und ihre Realisierungen. Die Archivnutzung bedeutet deshalb eine Gewinnung von Informationen aus den Spuren von vernetzten Aktionen. Sie benötigt eine ähnliche vernetzte Interpretationstechnik, wie sie das Internet von seinen Nutzern bei der Informationsgewinnung verlangt, wenn sie sich nicht in seinen unendlichen Verknüpfungen verlaufen wollen.

Die zunehmende Verbreitung des Internet fördert ein erneutes Erlernen des Denkens in Strukturen und Zusammenhängen, die lange Zeit durch die Leistungsfähigkeit von Datenbanken und darauf basierten Suchmaschinen sowie der Volltextrecherche in den Hintergrund geraten war. Doch die zunehmenden Mengen digital verfügbarer Angaben erhöhen die Trefferquoten in unüberschaubare Quantitäten und sie machen zudem die Begrenzung dieser Verfahren auf die Menge der von bereits einmal gewussten und abgespeicherten Informationen deutlich. Bei der Nutzung von Archivgut, bei der Lektüre von Findbüchern ebenso wie bei der Recherche im Internet werden gesuchte Informationen durch Analyse, Kombination und Interpretation der vorhandenen Angaben gewonnen, ohne dass sie bereits vorformuliert vorhanden sind. Vor allem die Glaubwürdigkeit und Verlässlichkeit dessen, was man sieht, muss man selbst einschätzen können. Die Techniken der Interpretation und der Absicherung von Annahmen werden mit der weiterhin zunehmenden Verbreitung des Internet zu allgemein erforderlichen Kulturtechniken moderner vernetzter Zivilisationen. Die Archive profitieren von der Ausbreitung dieser Techniken und methodischen Erfahrungen ihrer Benutzer und von deren zunehmender Fähigkeit, sich selbst ein Bild zu machen. Sie bieten heute schon mit ihren strukturierten Findmitteln die Möglichkeit

an, sich selbst davon zu überzeugen, welches Material vorhanden ist und festzustellen, was es aussagen kann und was nicht. Autonome Benutzer, die wissen, welche Fragen sie haben und sich selbst die Antworten erarbeiten, nutzen die Archive am besten. Ihnen öffnen sich die fremden Welten, die durch das Archivgut hindurch sichtbar werden. Deshalb liegt es im Interesse der Archive, die Autonomie ihrer Benutzer bestmöglich zu unterstützen. Die Wiederentdeckung der Strukturen fordert die Archive heraus, neue Wege und verbesserte Methoden zu ihrer Nutzung zu finden und bereit zu stellen. Sie bereitet ihnen gleichzeitig den Boden für neue, anspruchsvolle Präsentationen ihrer Bestände und die Verbesserung und Erleichterung des Zugangs zum Archivgut.

Anne van Camp: Integrating Access to Archives – The RLG Experience

What is RLG

RLG, (created as the Research Libraries Group, in 1974) is a non-profit consortium of over 160 research institutions from 15 different countries. Included are major universities, national libraries and archives, major museums, and other special kinds of historical collections. The mission of RLG is to improve access to information that supports research and learning through collaboration. The two major streams of activity in this regard include programs and projects that address problems shared by a number of institutions and the creation of information discovery and delivery services. Examples include the creation of new standards or best practices in the area of description of unique resources and the ability to bring together information about archival collections from diverse sources.

Historical Ways of Providing Access to Primary Materials

The history of archival descriptive practice is complicated and varies greatly from country to country. In North America, there is a patchwork of practice stemming from the history of decentralized custodial responsibility for historical records. National and state government archives have limited their responsibility to governmental records while special collections within research libraries and other independent historical repositories have tried to capture the personal papers, manuscript materials and other non-governmental records that document the North American experience. This approach to documenting culture relies heavily on altruism and subsidized support for the capture and preservation of records.

Descriptive practice reflects this historical development. Within government archives, records are described in a manner that reflects the individual agencies needs and preferences, often in a non-standard way. With the dispersed collection of non-governmental records, there is even more varied practice. Past methods of providing access to these unique research resources included: the published guides that some institutions were able to create; the unpublished guides available only in repositories; good reference staff; good scholarly guesswork; conference papers and footnotes in published works. Now that we have the means to make information about historical records accessible in electronic form there is a need to adhere to standards that allow this kind of information to be presented in a stable and sustainable form and to be shared across systems.

The history of standardized descriptive practice for special collections - even those within major universities - is a relatively short one. Until about 15 years ago special collections remained outside the mainstream of bibliographic descriptive practice. The belief that unique materials could not be described in a standard way was very strongly held by most archivists in North America and there was little guidance or leadership coming from any of our national institutions.

The development of the MARC-AMC format was at best a compromise that a few courageous institutions adopted in the early 1980's, embraced later by the profession after years of experimentation within certain communities and with a great deal of assistance from the institutions that encouraged the bibliographic integration of special materials alongside mainstream Library bibliographic information. Today in the RLG union catalog there are over 550,000 MARC-AMC records representing that many special collections - mostly from North American institutions.

In a 1992 report on historical documents called Using the Nation's Documentary Heritage, funded by the National Historic Publications and Records Commission and the American Council of Learned Societies, recommendations for creating better access to archives included training in use of sources, travel support for individuals to use sources, dissemination of research tools, finding aids and better access (this was of course prior to any internet availability.) And it included a criticism for not cataloging finding aids. Even with the ability to catalog at the collection level and include that information in a union database, much material remained inaccessible for a host of reasons. A survey conducted of RLG members in the early 1990's found that nearly 60% of special materials remained inaccessible because of lack of staff, money or expertise to make those collections known.

With the advances being made in the use of electronic technology, the world of special collections was still lagging behind other means of improved access to information. Clifford Lynch, Director of CNI, pointed this out very clearly when he said: "we need a new vision of opening up historically inaccessible special collections and linking them to both the existing and developing base of scholarly publication." (From an RLG Symposium in 1996, on Selecting Library and Archive Collections for Digital Reformatting.)

Current Efforts

As hard as it was to get the archival community to adopt the use of MARC-AMC for collection-level description, the development and application of the new standard for electronic encoding of full-text archival finding aids in EAD (Encoded Archival Description), has been astoundingly rapid. EAD gives the ability to provide full text finding aids in all their hierarchy in a web appropriate environment. This is a great leap

forward for enhancing access to collections using the preferred vehicle for information dissemination.

At the same time, there is much greater appreciation for the need to promote and integrate access to primary materials coming from the research community. A recent report from the Council on Library Resources and the American Council of Learned Societies (CLIR/ACLS) published in January 1999, on Scholarship, Instruction and Libraries at the Turn of the Century, had the following strategic recommendations:

- *Renewed emphasis on and encouragement of use of original materials in critical thinking,

- *Priority for developing finding aids for materials in all formats, along with a national standard for encoded archival description and a networked environment for them.

Another recent report, Where History Begins - a report on Historical Records Repositories in the US published by the Council of State Historical Records Coordinators (May 1998) gave a list of the most pressing problems confronting institutions, and Space and Storage were listed as number 1, Access and Finding Aids were listed as number 2.

Finding aids are the most important and effective tool for locating research materials, and improving access to them is critical. Encouragement of further sharing of these important resources is receiving a great deal of international attention as well. Web-based access mechanisms are maturing nicely and aggregating of this information is being done locally, regionally, nationally and internationally.

The availability of financial support for programs and projects that will make these special collections more accessible through electronic mechanisms has sparked a great deal of activity in this area. A few examples are the Online Archive of California (a state wide program to ag-

gregate finding aids from the California repositories), RLG's Archival Resources (an international program that provides unified access to dispersed collections,) and there are subject based projects like American Memory at the Library of Congress, and the Center for the History of Physics at the American Institute of Physics. The international MALVINE project sponsored by the European Union is another example of bringing together finding aids from several European countries. All of these projects are using the EAD standard so eventually they could all be integrated.

RLG's Archival Resources

RLG has built a service based on the use of EAD for archival finding aids and combined that with the ability to search across both these full-text documents as well as the 550,000 bibliographic collection level records that describe archival materials. The result is an immensely powerful tool for information discovery and delivery. There are today over 130 institutions contributing finding aids to the service and thousands of institutions represented in the bibliographic records. You can search and find related materials in archives from Minnesota to Australia to the Netherlands on subjects like technology transfer or global immigration. The topics are limitless and the discovery ability unprecedented.

This service has grown very rapidly and nearly 1000 new finding aids are being added each month when it is updated. Several efforts are underway to translate the EAD documentation so that it can be applied in French, Spanish, Portuguese, and more. The expectation is that growth will continue rapidly to make this a even more powerful resource for uniting dispersed collections and information and eventually this will be the discovery tool of choice for most archival researchers.

Continuing challenges

There is a great deal of tension between applying new technology and the perception that this is interfering with or preventing or even making worse the ability to carry out traditional duties such as processing, description, and preservation. I would argue these are not mutually exclusive but new tools for better description will increase the visibility of collections and contribute to greater access. Greater access is one of the highest priorities for cultural institutions and is a major criterion for measuring success. Successful programs will continue to grow and attract the support needed for carrying out all vital responsibilities. In other words, success breeds success.

There is also some controversy over providing unmediated access to information, removing the personal archivists as interpreters. This puts pressure on the creators of information to mediate carefully before putting information into the “unmediated” zone of the web. The whole purpose of this exercise is to make things easier and more accessible not more confusing than before. EAD sites that are not particularly easy to find, are not going to be found.

Remembering to design new access systems that are compatible with other access systems is also a concern. We cannot let users think that what is on the web is all there is and so these new systems must find ways of referring to those sources that are still not available in electronic form.

Uniting dispersed collections and overcoming language barriers are some of the most interesting challenges that I see. How can we identify and promote the linking of dispersed collections, perhaps through collaborative selection decisions, and overcoming barriers to truly international, interdisciplinary research.

Training for technical archivists is also a large challenge. Where are they being trained? Where will they come from? What will attract them to this work? These are questions the archival profession must deal with and with the growing demand for archivists with these new skills, the solutions must come quickly.

Finding aids are fast becoming locator tools for actual digital archival objects. Museums and other artifact collections are racing to the web. As our population ages and more people live longer, there are opportunities for lifelong learning that can be provided by knowledge providers. There is interest in finding new and better ways of introducing archival information into primary school curriculum. But many potential users will be new to us and perhaps our collection development will become driven by different demands for this information.

Greater use of these materials makes our collections of unique and special materials even more of an asset if we think beyond the immediate markets and mission we hold now. And those collections that are visible and known are the only ones that have a chance of being used in this new way.

For further information see the following web sites:

www.rlg.org (RLG home page)

www.rlg.org/arr/index.html (Information about the RLG Archival Resources service)

www.rlg.org/primary/index.html (Information about RLG's activities with the primary sources community)

www.rlg.org/primary/faprog.html (Information about RLG and EAD)

Kris Kiesling: The American Archival Community - Why We Need EAD

To understand the need for EAD in the United States, it is helpful to have a little background on the types of repositories prevalent in the country, learn a bit about the development of descriptive standards and how they are used by American archivists, and where Encoded Archival Description fits in that landscape.

Archival Repositories in the U.S.

The United States has a wide variety of types of archival institutions. There is of course the National Archives and Records Administration (NARA), which collects records from all branches of government and their offices, past presidents, legislators, and related materials. Each state also has a state archives, which may be part of a state library, state historical society, or an independent entity. Individual companies and organizations frequently maintain their own records over time, primarily for their own use but occasionally available to the general public. Many states, counties, and cities have historical societies that collect records of regional and local government offices, papers of individuals and families, and records of private organizations such as churches, civic groups, schools, and special interest groups. Colleges and universities typically have archives that document their history, and maintain records of college divisions and departments, student groups, and papers of faculty members. Frequently universities will also have a special collections department that collects rare books, manuscripts, and other cultural materials. Occasionally these special collections departments will also maintain the university records. Finally, there are independent manuscript repositories that have specific collecting interests, such as Shakespeare, early Americana, or other topics. Needless to say, there is a significant amount of overlap in the collecting focus of these institu-

tions, both topically and in the types of materials they receive. For example, it would not be unusual for a university to hold the papers of a U.S. senator who was an alumnus of that institution, or for a state historical society to hold the records of a multi-national corporation whose headquarters were in that state.

Descriptive Standards in the U.S.

In the past archivists have expressed various arguments against the use of descriptive standards. These include, "I don't need to share my data with anyone else," "Following standards takes too much time," and, my particular favorite, "My clients have special needs." Fortunately, one doesn't hear these arguments very much any longer, so clearly the attitude toward standards within the American archival community is changing.

This is not to say that American archivists have been entirely without standards. Collection descriptions¹ have traditionally taken two forms: catalog records and finding aids (inventories or registers). Archival catalog records consist of a brief summary description of an entire body of materials, usually containing index terms. In the past, many manuscript repositories created catalog records for individual items in a collection, a practice that has almost entirely disappeared. Finding aids are multilevel descriptions that contain progressively more detailed description of the entire body of materials through descriptions of individual series (subgroupings within the collection reflecting format of material, function, or activity), subseries, and finally files or items. This is a very different system than the German registratur, which focuses on description of a file

¹ Throughout this paper I will use the word "collections" to mean all types of materials collected by a repository, including institutional records, fonds, personal papers, and artificial gatherings of materials.

that represents an individual transaction. In the U.S. the content of catalog records and finding aids was roughly the same from repository to repository, but there was little in the way of rules or guidelines to be followed until the latter part of the 20th century.

From these humble, paper-based beginnings archivists quickly embraced the opportunities the online world had to offer. In the mid-1980s, a special MARC (MACHine Readable Cataloging) format known as Archives and Manuscripts Control (AMC) was developed for archival materials, enabling archivists to place collection-level descriptions of their holdings in online catalogs along with “traditional” library materials. In order to do this archivists adopted bibliographic descriptive standards such as the *Anglo-American Cataloging Rules*, 2nd ed. (AACR2), and began to use controlled vocabularies used by the library world, including the *Library of Congress Subject Headings* (LCSH), and the *Art and Architecture Thesaurus* (AAT).² In the early 1990s archivists almost immediately recognized the advantages of putting finding aids on Gopher systems, the early predecessors of the Internet. Now we have progressed to the point where most American repositories have finding aids available on the Web, either in HTML or Encoded Archival Description (EAD).

The ability to make collection descriptions available to a much wider audience (potentially the entire world) than had been previously possible has had the effect of encouraging American archivists to develop and use descriptive standards. Visibility equals scrutiny. Descriptions that

² *Anglo-American Cataloging Rules*, 2d ed., rev. Chicago: American Library Association, 1988.

Library of Congress Subject Headings, 19th ed. Washington, DC: Cataloging Distribution Service, Library of Congress, 1996. Prepared by the Cataloging Policy and Support Office, Library Services.

Art & Architecture Thesaurus, 2d ed. New York: Oxford University Press, 1994.

had seemed adequate when available only within the repository's walls, where a reference archivist was present to explain the finding aids as well as the collection holdings, now are seen in the cold light of the Internet where there is no reference archivist and the finding aids are available not only to experienced researchers but also to school children and the general public. Systems have developed to the point where it is possible for a researcher to begin his or her work by searching a university or research library's online catalog, find records for archival materials along with books, maps, and sound recordings relevant to the search topic, click on a link in the summary catalog record for the archival materials and bring up the detailed finding aid for those materials. In some cases, digitized images of the materials in the collection are available through links from the finding aid. These are definitely developments around which to build descriptive standards!

Types of Descriptive Standards

There are a number of types of descriptive standards that American archivists currently use with regularity. These include data structure standards, data content standards, and data value standards.

Data structure standards define elements of information and, in many cases, prescribe the order in which they must appear and whether individual elements are repeatable. Structure standards enable data communication and exchange, most often via computer. Examples of data structure standards include MARC21³ for online catalog records, the General International Archival Description (ISAD(G)) for all types of descriptions, and Encoded Archival Description for finding aids.

³ In 1999 the U.S. and Canadian versions of the MARC format (USMARC and CAN/MARC) were combined to form MARC21.

Data content standards comprise the rules to apply when formulating the information that is placed in the data elements. For example content standards provide instructions on how one expresses, in a consistent manner, the number of pages in a book or the physical extent of an archival collection. In the U.S., there are a number of content standards that archivists use, including the aforementioned AACR2; *Archives, Personal Papers, and Manuscripts* (APPM), a set of general cataloging rules for archival materials; *Graphic Materials* (GM), which applies to photographs, posters, and other visual media; *Archival Moving Image Materials* (AMIM), which is used for cataloging films and videotapes; and the *Oral History Cataloging Manual*, to name a few.⁴ APPM and the more specific media cataloging rules were the result of dissatisfaction with AACR2's treatment of archival materials. The *EAD Application Guidelines*⁵ provide some guidance in the creation of the content of finding aids, but focus more on the proper use of the EAD elements. There is a new initiative, funded by the U.S. National Endowment for the Humanities, to create a joint Canadian/U.S. content standard that can be used to create catalog records, finding aids, or other types of descriptions of

⁴ Hensen, Steven L. *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries*, 2d ed. Chicago: Society of American Archivists, 1989.

Betz, Elizabeth W. *Graphic Materials: Rules for Describing Original Items and Historical Collections*. Washington, DC: Library of Congress, 1982.

White-Hensen, Wendy. *Archival Moving Image Materials: A Cataloging Manual*. Washington, DC: Motion Picture, Broadcasting, and Recorded Sound Division, Library of Congress, 1984.

Matters, Marion E. *Oral History Cataloging Manual*. Chicago: Society of American Archivists, 1995.

⁵ *Encoded Archival Description Application Guidelines, Version 1.0*. Chicago: Society of American Archivists, 1999. Also available online at <http://lcweb.loc.gov/ead/>.

archival materials. Completion of this content standard is scheduled for late 2002, and in the U.S. it will replace APPM.

Data value standards include code lists and thesauri—controlled vocabularies that enhance retrieval of the data, particularly in online systems. The value standards used by U.S. archivists are borrowed for the most part from the library world; the advantages of using them are obvious if archival catalog records are being added to library catalogs, but also apply to online finding aids. These standards include the already mentioned *Art and Architecture Thesaurus* and *Library of Congress Subject Headings*, as well as the *Library of Congress Name Authority File* (LCNAF), the LC *Thesaurus for Graphic Materials* (LCTGM), and MARC code lists for languages, countries, and occupations, and many other resources.⁶

Archivists in the U.S. actively participate in a number of standards organizations, nationally and internationally. Within the Society of American Archivists, a Standards Committee monitors relevant standards being developed and used in the library, archives, and museum communities. These not only encompass descriptive standards, but include preservation, environmental, electronic record, and digital object standards as well. The Standards Committee makes recommendations to the SAA Council regarding endorsement of a particular standard for use by SAA members. The Standards Committee has a Technical Subcommittee on Descriptive Standards which deals entirely with standards relating to archival cataloging, finding aids, metadata, and other areas affecting ar-

⁶ *Thesaurus for Graphic Materials*. Compiled by the Prints and Photographs Division, Library of Congress. Washington, DC: Cataloging Distribution Service, Library of Congress, 1995.

The various MARC code lists are available online at <http://lcweb.loc.gov/marc/>.

The LC Name Authority File is available by subscription via the OCLC or RLIN bibliographic networks.

chival description. The EAD Working Group reports to the Technical Subcommittee. SAA also has a liaison to MARBI (Machine-Readable Bibliographic Information) committee within the American Library Association. MARBI is responsible for the continued development of the MARC21 format for cataloging. SAA also is a voting member of the National Information Standards Organization (NISO) and comments via NISO on standards being considered by the International Standards Organization (ISO). Two U.S. archivists currently are members of the Committee on Descriptive Standards of the International Council on Archives (ICA).

Encoded Archival Description

Where does EAD fit in the larger standards picture? There had been tacit agreement among U.S. archivists regarding what constitutes a “good” finding aid (and, in fact, many repositories followed the models established by the National Archives and by the Manuscripts Division in the Library of Congress), but there was not absolute consistency between repositories in the structure or even the content of their finding aids. “Local practice” was the norm, and was potentially defensible when all access was within the repository walls. But once finding aids could be placed on the Web for universal access, it became apparent that perhaps something more rigorous was needed. Here was an excellent opportunity to establish a data structure standard for finding aids, whether in paper or online form, as well as an opportunity to standardize the presentation of finding aids on the Web, making it easier for remote users to understand and get access to the materials.

In the earliest days of EAD development, during the Berkeley Finding Aid Project, functional requirements for an encoding standard for finding aids were established:

- It must be an open standard (i.e. non-proprietary and available to anyone)
- It must be platform independent (usable on a Macintosh, Windows, or UNIX based operating system)
- It must support long-term data storage and migration
- It must represent the hierarchical relationships inherent in archival materials
- It must support navigation and retrieval in an online system

It was immediately clear that HTML and MARC would not fulfill the functional requirements, so Standard Generalized Markup Language (SGML) was chosen for the Berkeley project.

EAD is a data structure standard. It establishes data fields into which information is placed, and controls the order in which those fields can be used (albeit a very flexible order) and whether or not the fields can be repeated. EAD is inherently multilevel, and can be used for descriptions of archival materials from the collection to the item level, and anything in between. It is compatible with ISAD(G),⁷ and, like ISAD(G), is intended to be used with national descriptive rules and guidelines (content standards).

Most importantly, however, EAD provides an opportunity for archivists (and not just those in the U.S.) to rethink and improve their descriptive practices. Many repositories, in converting their older finding aids to EAD, have chosen to “upgrade” the content and structure of those leg-

⁷ The first version of EAD, released in 1998, was compatible with the first edition of ISAD(G). The second edition of ISAD(G) was published in 2000, and the next version of EAD, to be released in early 2002, will be reconciled. ISAD(G) is available at <http://www.ica.org>.

acy files to take advantage of what EAD has to offer, keeping in mind the needs of the end users in unmediated cyberspace with no reference archivist to guide them. EAD has already proved itself to be a mechanism for international data exchange. Archival Resources, a service of the Research Libraries Group, gathers finding aids from repositories in several countries and makes them available in a single database. Other multi-institution sites of EAD-encoded finding aids spring up on a fairly regular basis. EAD is also a catalyst for the creation of international content standards for finding aids, such as the Canadian/U.S. project mentioned above. It is hoped that the creation of such content standards will make descriptions of archival materials intelligible to the entire potential Web audience, not just experienced researchers and other archivists.

Catherine Dhérent: French experiences with the adaptation of EAD

In 1996, John Haeger, vice-president of the Research Libraries Group (RLG) proposed to Mr Erlande-Brandenburg, Director of the Archives de France, to send an archivist to America to see what was made at that time in the electronic field in Records centres and libraries. It was in this context I discovered the DTD (EAD) the first time. I was very interested by the use of this new tool I never heard of before, and I saw what was made with it in Stanford, Berkeley libraries, Minnesota historical society... I was deeply impressed by it. And once returned in France, I made an enthusiastic report on the works realised with this tool and its interest to publish our finding aids on the web.

But at that time, none of the Records centres in France, even the five centres of National Archives, had web pages. In French Archives, nobody did know exactly what was internet, e-mail and so on. Very few people knew ISAD(G) in details and respect it literally. In the Parisian centre of National Archives, there were more than 60 different data bases conceived with different electronic tools and without an overview and a policy of a standardized and rationalized communication to the public.

Moreover, EAD was produced by American librarians and in France, archivists thought that an American tool for description could not be effective for French practices. And in France, we began to standardize very late. Therefore, my report was not understood and was put in a drawer. It arrived a little too early!

But a few months later, things changed for some reasons. At first, France is a centralized country and it is a chance in this issue. In fact, the first reason of the change is that the French government as others almost everywhere in the world, settled little by little the Information

Society and elaborated a program for this aim (PAGSI). It was decided to digitalize the most important documents of the country and the Ministry of the Culture had to assume it. In this ministry, there is a special office, a mission which is dealing with all subjects of Research and Technology (Mission de la Recherche et de la Technologie). It is responsible for the setting up of the governmental plan in patrimonial institutions. This plan have profited by substantial funds since the beginning and has allowed even small local institutions to digitalize their collections, essentially birth, marriage and death acts, that is called « état civil » in French, maps particularly those of the Napoleonian cadastre realized between 1820 and 1835 and others documents, precious or useful to the researchers and often requested by them.

In this plan of digitization of the most important documents, national institutions, for example National Archives, had the first place at the beginning. The first pictures of French archival documents on a web site was put in the base called ARCHIM, on the web pages of the ministry of the Culture. It was in 1997. Since this time, a curator has been in charge of the base and she has paid a great attention to all the technological developments of the tools.

The Ministry of the Culture is promoting the use of free and standardized electronic tools and accordingly of XML, particularly since 1998. At the same time, the premier created an interministerial office to help all administrations to choose the good technologies, the most standardized, the less expensive on the long-term, and the most durable. It is the MTIC, acronym for Mission interministerial de soutien à l'introduction des technologies de l'information et de la communication dans l'administration (Interministerial agency to help administration to use technologies of information and communication). This office is maintaining a web site on which there are forums on technologies and specially on free soft-

wares, on XML and it organizes training, information days on these subjects.

In this context, the Ministry of the Culture is, in the French government, one of the most attentive to XML and to the different types of structured documents, to the DTDs. Its goal is to help each patrimonial institutions under its control to exchange very heterogeneous data and to build together new databases. So it searched which DTD existed yet and when they did not exist, it was decided to build new DTDs. Thus for some institutions, DTDs have been elaborated : for example for inventory of Historical monuments or for the catalogue of printed books. The first step was some years ago an European project, Aquarelle.

For archives, a DTD was existing. It was EAD. So the ministry proposed to the Direction des Archives de France, to test this DTD. At that time, French curators were not entirely convinced of the relevance of an American tool : they thought that American and French archival practices were extremely different. And they did not see what EAD could add to ISAD (G). They did not see the difference between the rules for the description of the content and the electronic tool to implement these rules. So they began the test without great enthusiasm.

A finding aid of a local records centre was chosen. The Direction des Archives de France wanted to experiment EAD with a complex inventory with many levels. It must also be a recent one, well written, consistent with ISAD (G). A floppy disk with the Word document was also necessary. This finding aid was this of the series M of the department of Pyrénées-Orientales, which contains all the fonds of the general administrations in a French local institution, a "department", from 1800 to 1940.

The result was greatly better than expected. We used easily EAD with nine levels and few difficulties. Generally, all the conversion of the data

was automatically realized into EAD and the problems did not come from the EAD but from the original finding aid. Even after having read it carefully, we did not see some anomalies, lacks in the standardization, for example in defining the producers of the documents, because in this finding aid there were several fonds.

But this test was not exemplary because at that time we did not have enough knowledge of EAD. Neither did we ask our vendor to use stylesheets or research tools. We wanted only to check the compatibility of the structure and of the elements of EAD with our practice.

We realize at this occasion how much constraining, misleading and confusing was our use of French series. They are not organic series as in ISAD (G) but only a convenience for physical arrangement and classification. They belong to an old classification plan (1841) not always well adapted today. The French series appear at the beginning of all inventories, at the first place, in the shape of a letter. For example in all archives of a departmental institution, the series O contains all the fonds produced by the offices in charge of the local buildings before 1940, the series R all the fonds on the army and defence before 1940, the series U all the fonds of the law courts before 1958 and so that. For some curators the series were become little by little more important than the fonds and often today the series are hiding the fonds and some fonds are mixed up in a series.

Therefore, since this test, we are telling in each training and information days that the series are just a convenience and that the most important thing is to highlight the fonds and to make this notion more understandable to our readers.

But at the time of this test, we did not wonder if we had to create only one instance of EAD or as many instances as they were fonds in this finding aid. And our encoder, a specialized company, chose an attribute

level « otherlevel » « série » for French series of the classification plan. It was a little confusing for non-French readers.

This test was finished in 1999. At that time, we heard of an European meeting on EAD in London in the middle of October. We asked to participate to it and were impressed by the works of some Records centres or libraries, for example PRO, Swedish and Finnish patrimonial institutions. But we were also convinced that the language of all the documentation and training was a problem for French curators. So we decided to create a French speaking working group. This one was settled at the end of January 2000. We had a meeting with 40 French speaking colleagues, most of them from France and some from Switzerland, Germany (Frauke Laufhuetten from the Bundesarchiv in Coblenz), Spain. We decided to translate the Tag Library and to issue an electronic newspaper.

This bulletin is put on line every two or three months since March 2000. It gives brief information on the different working group on EAD, on tests and projects, on training sessions... Several hundreds of people are now inscribed on the list to receive this little news. It is at the url : <http://www.archivesdefrance.culture.gouv.fr/fr/publications/DAFbulldtd.htm>

To translate the Tag Library and after it, the Application Guidelines, we set up a group inside the French society of standardization, AFNOR, in April 2000. We are about fifteen persons coming from archives, libraries, museums, private companies and university. We meet once per month and in the meantime, we work in little groups of two or three persons to prepare the translations. The texts are thus at the stage of Translated. During the monthly meeting, we examine these translations and move them to the stage of Revised. We have yet translated the Overview, Design Principles, a great part of Attributes and nearly all the

elements and we have moved the third part of them to the stage of Revised. The examples given in the description of each element of the Tag Library will be French.

This translation is a mean for us to check the compatibility of EAD with the practices of French archives, libraries and museums. We hope that the first version of our French translation will be achieved at the end of this year. We have decided to put it on line on the websites of Direction des Archives de France and of National Library. The access will be free. The Society of American Archivists has accepted this, we are setting an information on the conditions of the translation, on the rights of SAA and on its agreement.

In October 2000, the Society of American Archivists suggested France becoming one of its international members. Since the Direction des Archives de France is representing our country near the Society. In this context, we sent some submissions as others members for the future new version of EAD, submissions which were discussed in Washington one week ago.

At the same time, a new test was decided : the encoding of the "Etat général des fonds" that is meaning general summary of records of public archives (vol. 2 and 5) and Etat général des inventaires or general repertory of finding aids of these same records. The first one consists in two volumes printed at different dates, the first volume for the records of the period 1789-1940 preserved in Paris and the second one for the records of the period 1940-1958 preserved in the same Archives : they were computerized with OCR (optical character recognition) in 1998. Its structure is simple and a hierarchy of information is apparent (French series, fonds, sub-fonds, series...). The description is output into columns, one for the identification number or a range of identification numbers, one for the analysis or title, one for the dates. So it appeared well

adapted to the EAD at first. But there are nevertheless some difficulties. The vol. 5 of the summary of the fonds contains information on the finding aids, therefore a part of the repertory of all the finding aids of National Archives. The two volumes of the summary of records (Etat des fonds) which must be mixed, have some redundancies. Sometimes there are also several levels in the same description, without explicit remarks and sometimes the characteristics of the printed format do not help to define the level of the description. And so that. Thus it was decided that before and after the most part of the work having been made automatically, National Archives will spend some time to clean the original files and the results little by little. Nevertheless, every issue will be solved rather easily. This test will be finished probably during the Summer.

Another test was decided but did not yet begin. It will be applied to the special finding aids used for the contemporary documents regularly transferred from a public institution to the Archives centre. This finding aid is called in French : *bordereau de versement* or in English transfer list. A transfer list is written by the office which transfers the documents, office which can also be the producer. After that the transfer list is checked by Archives centre and finally, the documents are received by the Archives. This kind of finding aid is particular. It is a part of a repertory of a fonds that is not yet exactly known. It can also contain several parts of several fonds. These finding aids are used for non-closed fonds.

So we want to check if the structure of an EAD instance can be easily modified and completed little by little, as and when required, every time that a new transfer of records of the same fonds is made to the Records centre. It could also be interesting also to see if it would be possible to share the work as it is the case today manually, and to let the producer

office fill some elements of the EAD instance before the Records centre checks it and completes it.

During all this program of tests and experiments, the Direction des Archives de France was granted by an American foundation, Delmas Grièble Foundation (from New-York) thanks to the help of RLG. Four finding aids are now being encoded with this subsidy. The work will be achieved and put on our web site and on the website of the RLG at the end of May or at the beginning of June. Two of them are completing the test on the summary of records as it is applied to a part of this summary dedicated to the part of private archives. They are the structured list of the private fonds of Historical centre of National Archives (Paris) and the same of the National Archives devoted to the archives of private companies, located in Roubaix in the North of France. These lists are compatible with the international standard of description ISAD (G).

One another of these finding aids is the inventory of Napoleon archives, fonds of great interest, known by American researchers and a lot of documents of it are microfilmed and are being digitalized. So the DTD element `<dao>` which permits to link another electronic file, for example pictures file with the EAD instance, will be used for that. It will be the first time during our tests.

Finally, we have decided to propose to the Delmas foundation a database conceived by a local Records centre, the Archives of the department of Aube (this area is at 150 kilometres on the east of Paris and has prestigious fonds as this of Clairvaux Abbey). The encoding of the database of this fonds was also agreed by Delmas foundation. And its encoding is rather easy.

The ministry of the Culture has also decided to implement a XML technical platform on the web with tools of production, management and diffusion of electronic documents. The specification has just been writ-

ten in April. All tools must be based on free software. Everyone could use these tools. So now I recommend to our archivists not to acquire software for the archival description but to wait some months yet and to use these free XML and EAD tools. There will be a module to transfer databases easily in XML software. The ways of edition will be internet or intranet, e-files or paper, as everyone wants. The requirements of the specification are easiness for the users and for the checkers, possibility to integrate digitalized pictures, electronic bibliographical or geographical files, possibility to share a work and the writing of a file also with people working on laptops for example for the inventory of Historical monuments, reusability, respect of the requirement of the W3C and so that.

So the Direction des Archives de France is greatly helped in its work by its Ministry.

Some other French institutions wish today to use EAD soon and have relationship with Archives de France :

The new National Institute for art history (INHA) for the output of the inventories of archives of artists and galleries, on website.

The library of Lyon for writers, editors, art galleries archives.

The National Library (BnF) for special collections and manuscripts.

The society of French librarians (ABF) is thinking also about EAD and promote information conference to its members on DTD, TEI or EAD.

Some have yet made experiments : the Institute Memory of contemporary edition (IMEC) is a member of the European Malvine project on the contemporary manuscripts. A curator of the Napoleon Foundation has realised an inventory of the documents printed by the Conseil d'Etat (State Council) at the beginning of the 19th century. The finding aid is on the website <http://www.napoleonica.org/>. It is not a pure encoding in

EAD and not automatically realised because this curator did not have the tools to do that but the result is rather good.

Therefore now, the context in France seems to me good for the development of these new practices. I think that EAD can help French archivists to recover the good methods and to respect the requirement of the ISAD (G), going from the general to the particular, and structuring the fonds from the high level to the item level. Since two decades, the use of simpler databases software, non oriented-object, lead French curators to forget the principle of provenance, the respect of fonds which at the beginning, in the middle of the nineteenth century, was a French idea. The French series of the national classification schemes have little by little hidden the fonds. Many of us have described records without logical classification, as they find them by chance on the shelves, box after box, file after file. It was the consequence of the accession of too many records produced by an administration which do not register them as it is the case in Germany.

I am sure that XML and EAD are today the most effective tools able to help archivist in their work. It is also the best way to strengthen and reinforce the net of French Archives. 800 Archives are under the control of Direction des Archives de France but didn't have a common tool to communicate the content of their collections to the public. EAD can be now this tool. It is the reason why since two years I am trying to promote it in our country, setting up meetings, training and education. For example, we will have a workshop managed by the Society of American Archivists and a conference given by Kris Kiesling from the Texas university, in the background of the international training set up by Direction des Archives de France at the end of May. Some lessons on EAD are now also taught to the students of the National School of Patrimony and will be developed in the future.

Thus, with little devices, few money and few people, we are trying to implement and develop in our country the standardization and electronic tools conceived by American archivists and we are happy to contribute a little to their improvement.

Bernhard Grau: Das deutsch-amerikanische Projekt "Gemeinsames Fachkonzept Online-Erschließung" - Zur Übertragbarkeit der Encoded Archival Description (EAD) auf die archivische Praxis in Deutschland¹

I.

Mit seiner zunehmenden Ausbreitung gewinnt das Internet auch für die deutschen Archive eine ständig wachsende Bedeutung. Die Einrichtung eigener Homepages mit einem möglichst umfassenden Informationsangebot gilt fast schon als Selbstverständlichkeit. In Verbindung damit steigt auch die Erwartung, die archivischen Findmittel via Internet abfragen zu können. Vielfach wird sogar gefordert, die Archivalien selbst online zugänglich zu machen. Lässt sich letzteres derzeit wohl nur in Form einer qualifizierten Auswahl besonders wichtiger Dokumente verwirklichen, so besitzt die Verfügbarmachung von Erschließungsinformationen über das World Wide Web schon jetzt eine realistische Perspektive.²

Im deutschsprachigen Raum hält sich das Angebot von Online-Findmitteln allerdings nach wie vor in überschaubaren Grenzen. Dabei spielt mit Sicherheit eine Rolle, dass die Verzeichnungsergebnisse bislang nur teilweise in digitaler Form vorliegen, ja dass oft gerade diejeni-

¹ Der auf dem Marburger Kolloquium gehaltene Vortrag wurde für die Drucklegung leicht überarbeitet, um die Ergebnisse des zweiten Treffens mit der amerikanischen Arbeitsgruppe zumindest partiell noch berücksichtigen zu können.

² Siehe hierzu Karsten Uhde, Archive und Internet, in: Der Archivar, 49. Jg. (1996), Sp. 205-216; Paul Flamme, Udo Herkert, Volker Viergutz, Hinweise zur Darstellung staatlicher Archive und Archivverwaltungen im WorldWideWeb des Internet, in: Der Archivar, 51. Jg. (1998), S. 218-228; Mechthild Black-Veldtrup, Erschließung im Umbruch, in: Der Archivar, 51. Jg. (1998), S. 608-618; Mario Glauert, Anforderungen an eine Online-Beständeübersicht und eine archivische Homepage (2001) = <<http://www.uni-marburg.de/archivschule/fv5.html>>.

gen Bestände, die sich für eine Präsentation im Internet besonders gut eignen würden, nur durch analoge Hilfsmittel wie Findbücher oder Karten erschlossen sind. Die Retrodigitalisierung verspricht hier Abhilfe, ist aber wegen der heterogenen Form und Struktur der traditionellen Findmittel eine zeitintensive Aufgabe. Hinzu kommt das Fehlen einheitlicher Standards für die Erstellung und die Online-Präsentation digitaler Findbücher. Entsprechende Initiativen bringen daher für das einzelne Archiv nicht nur technische, sondern auch eine Reihe methodischer und praktischer Herausforderungen mit sich.³

Anders stellt sich die Situation für den anglo-amerikanischen Sprachraum und hier vor allem für die Vereinigten Staaten von Amerika dar. So hat die Zahl der von den amerikanischen Archiven angebotenen Online-Findbücher bereits einen beachtlichen Umfang erreicht. Zudem gibt es zentrale Eingangs- und Nachweisportale, die eine Institutionen übergreifende Recherche ermöglichen und damit die Suche nach einschlägigen Archivalien spürbar erleichtern. In Form der Encoded Archival Description, einer einer auf die Beschreibung archivischer Materialien zugeschnittenen Document Type Definition (DTD), liegt auch schon ein standardisiertes Konzept für die Erzeugung onlinefähiger Findmittel vor. Mit Erscheinen der Version 1.0 hat die EAD im Jahr 1998 das Probestadium

³ Siehe hierzu Angelika Menne-Haritz, Online-fähige Repertorien? Einige Überlegungen zur Interaktivität von Archivfindmitteln, in: *Der Archivar*, 49. Jg. (1996), S. 604-610; Mechthild Black-Veldtrup, Findbücher im Internet - Möglichkeiten ihrer Präsentation, in: Angelika Menne-Haritz (Hrsg.), *Archivische Erschließung: Methodische Aspekte einer Fachkompetenz. Beiträge des 3. Archivwissenschaftlichen Kolloquiums der Archivschule Marburg, Marburg 1999* (Veröffentlichungen der Archivschule Marburg, Nr. 30), S. 123-138; Edgar Büttner, Internet-Präsentation von Online-Findmitteln - Voraussetzungen und Bestrebungen des Bundesarchivs, in: *Der Archivar*, 54. Jg. (2001), S. 212-215.

hinter sich gelassen. Sie stellt zweifellos einen innovativen Lösungsansatz dar, der eine ganze Fülle von Vorteilen bietet.⁴

Allein dieser Befund gibt ausreichend Anlass, sich intensiver mit der Erschließungspraxis in den USA auseinanderzusetzen. Entsprechende Pläne wurden auch bereits in Zusammenhang mit dem von der Deutschen Forschungsgemeinschaft geförderten USA-Besuch einer Gruppe deutscher Archivare im Jahr 1999 entwickelt.⁵ Im Februar des Jahres 2000 wurde dann von der Leiterin der Archivschule Marburg, Angelika Menne-Haritz, bei der DFG der Antrag auf Förderung eines deutsch-amerikanischen Gemeinschaftsprojekts gestellt, dessen Ziel es sein sollte, ein gemeinsames Fachkonzept für die Online-Präsentation von archivistischen Findmitteln zu erarbeiten. Auf diese Weise sollte vermieden werden, dass sich auf internationaler Ebene unterschiedliche Standards herausbilden, die zu Schwierigkeiten bei der grenzüberschreitenden Zu-

⁴ Jackie M. Dooley (Hrsg.), *Encoded Archival Description: Context, Theory, and Case Studies*, Chicago 1998; *Archives and Museum Informatics*, Volume 12 (1998), S. 169-305 (= Richard Rinehart, Anne-Gilliland-Swetland (Hrsg.), *EAD in Action: Applications of the Encoded Archival Description*); Daniel Pitti, *Encoded Archival Description. An Introduction and Overview*, in: *D-Lib Magazine*, Vol. 5, Number 11, November 1999 <<http://www.dlib.org/Dlib/november99/11pitti.html>>; Wendy M. Duff, Kent M. Haworth, *Advancing Archival Description: A Model für Rationalising North American Descriptive Standards*, in: *Archives and Manuscripts*, Vol. 25, S. 194-217, hier insbesondere S. 199-201. Zur Entwicklungsgeschichte der EAD siehe auch <http://sunsite.berkeley.edu/FindingAids/EAD/history.html>. Zur Genese der US-amerikanischen Erschließungsstandards allgemein siehe *Standards for Archival Description. Reports of the Working Group on Standards for Archival Description*, in: *American Archivist*, Vol. 52 (1989), S. 440-461.

⁵ Sigrun Eckelmann, Hans-Dieter Kreikamp, Angelika Menne-Haritz, Wilfried Reinhaus, *Neue Medien im Archiv: Onlinezugang und elektronische Unterlagen. Bericht über eine Studienreise nach Nordamerika, 10.-21. Mai 1999, Marburg 2000* (Veröffentlichungen der Archivschule Marburg, Nr. 32).

sammenarbeit und beim Austausch von Erschließungsinformationen führen könnten.

Die Genehmigung des Projekts durch die DFG erfolgte bereits im April des Jahres 2000. Daraufhin konstituierten sich eine deutsche und eine amerikanische Arbeitsgruppe, die sich aus acht beziehungsweise sieben Personen zusammensetzen. Während in der deutschen Arbeitsgruppe Kolleginnen und Kollegen aus dem Bundesarchiv, den Landesarchivverwaltungen und dem kommunalen Bereich vertreten sind, wird die amerikanische Seite vor allem von Archivarinnen und Archivaren aus dem staatlichen, dem universitären und dem bibliothekarischen Bereich repräsentiert. Ein Teil der Mitglieder gehört zugleich der für die Pflege der EAD zuständigen Working Group an.⁶

Der finanzielle Aufwand der amerikanischen Arbeitsgruppe wird vom Council on Library and Information Resources (CLIR) bestritten. Dabei handelt es sich um eine private Einrichtung, die es sich zum Ziel gesetzt

⁶ Die Mitglieder der amerikanischen Arbeitsgruppe sind: Jackie Dooley (Leiterin des Archivs der Universität von Kalifornien, Irvine), Michael Fox (Direktor der Minnesota Historical Society), Steve Hensen (Direktor der Abteilung "special collections" der Duke Universität, Durham/North Carolina), Kris Kiesling (Leiterin des Archivs des Humanities Research Centers der Universität von Texas, Austin), Kathleen Roe (Direktorin des Staatsarchivs und der Schriftgutverwaltung des Staates New York), Abby Smith (Vertreterin des Council on Library and Information Resources / CLIR), Richard Szary (Leiter des Universitätsarchivs der Yale-Universität, New Haven). Am Treffen in Washington nahm außerdem die Präsidentin des CLIR, Deanna Marcum, teil. Die deutsche Arbeitsgruppe setzt sich zusammen aus: Dr. Nicole Bickhoff (Landesarchivdirektion Stuttgart), Dr. Mechthild Black-Veldtrup (Hauptstaatsarchiv Stuttgart), Dr. Edgar Büttner (Bundesarchiv Koblenz), Beate Friedrich (Bundesarchiv, Stiftung der Parteien und Massenorganisationen der DDR, Berlin), Dr. Bernhard Grau (Bayerisches Hauptstaatsarchiv), PD Dr. Angelika Menne-Haritz (Archivschule Marburg), Klaus Tempel (Geheimes Staatsarchiv Preußischer Kulturbesitz, Berlin), Katharina Tiemann (Westfälisches Archivamt Münster). Die aktuelle Liste der Mitglieder der EAD-Working-Group findet sich unter <<http://lcweb.loc.gov/ead/eadwg.html>>.

hat, Projekte und Forschungsvorhaben anzustoßen und zu fördern, die sich um einen verbesserten Zugang zu bibliothekarischen und archivischen Materialien bemühen. Ein wichtiger Aspekt ihrer Tätigkeit ist der Anspruch, Lösungen zu fördern, die auch international tragfähig und verwendbar sind.⁷

Nach einer Vorbereitungsphase, die auf deutscher Seite zur Formulierung von Zielvorstellungen für die verschiedenen Aspekte der Problematik genutzt wurde, fand am 23. und 24. Juni 2000 in Washington das erste von insgesamt zwei Treffen mit der amerikanischen Seite statt. Zum Abschlusstreffen besuchten die Vertreter der USA am 14. und 15. Mai 2001 die Archivschule in Marburg. Bei diesen Zusammenkünften kam es zu äußerst intensiven Diskussionen, die zunächst dem Abgleich von Unterschieden in der Praxis der US-amerikanischen und der bundesdeutschen Archive dienten. Dabei wurde auch deutlich, dass der fortgeschrittene Entwicklungsstand der Encoded Archival Description eine Überarbeitung und Anpassung an die Anforderungen der deutschen Archive wohl nur noch in gewissen Grenzen zulässt. Der deutschen Seite musste es daher vor allem darum zu tun sein, die Vereinbarkeit der EAD mit den in Deutschland gebräuchlichen fachlichen Standards und Arbeitsmethoden zu prüfen und gegebenenfalls Möglichkeiten zur Anpassung an die eigenen Anforderungen zu ermitteln.

II.

Bei der Beschäftigung mit der Encoded Archival Description sind zunächst die großen Vorteile zu beachten, die sie bietet. An erster Stelle ist dabei auf ihren großen Verbreitungsgrad hinzuweisen. In den USA

⁷ <<http://www.clir.org/>>.

stellt sie bereits so etwas wie einen Standard dar, der in einer Vielzahl von Institutionen aus dem Archiv-, Bibliotheks- und Museumsbereich zur Anwendung kommt. Auch auf internationaler Ebene ist eine rasche Verbreitung festzustellen. Nicht nur in Kanada und in Großbritannien kommt die EAD zum Einsatz, auch in Australien, Spanien, Frankreich und Griechenland sind entsprechende Aktivitäten im Gange.⁸ Dies zeigt nicht nur, dass es sich bei der EAD um ein praxistaugliches Produkt handelt, sondern lässt zugleich erwarten, dass sie sich auch international zu einem Quasi-Standard entwickeln wird.

Der Erfolg der EAD ist nicht zuletzt durch die Möglichkeiten bedingt, die sie dem Anwender schon jetzt eröffnet. So ist nicht zu übersehen, dass es sich dabei um ein universell einsetzbares und vergleichsweise zukunftssicheres Produkt handelt. Dafür sorgt zunächst die Tatsache, dass die EAD auf dem SGML-Standard aufbaut und mittlerweile auch auf XML-Basis verfügbar ist. Diese Tatsache bietet Gewähr dafür, dass Findmittel, die auf dieser Grundlage erzeugt werden, auch längerfristig les- und verwendbar bleiben. Hinzu kommt, dass das EAD-Konzept gut dokumentiert und ohne Einschränkungen frei zugänglich ist.⁹ Software-

⁸ Siehe etwa Daniel Pitti, Encoded Archival Description. An Introduction and Overview, in: D-Lib Magazine, Vol. 5, Number 11, November 1999 = <<http://www.dlib.org/Dlib/november99/11pitti.html>>; Bulletin des Archives de France sur la DTD (EAD) No. 1 -8 (= <<http://www.archivesdefrance.culture.gouv.fr/fr/publications/index.html>>); George N. Mitrofanis, Encoded Archival Description (EAD) and the dissemination of historical information: the application of EAD in the National Bank of Greece Historical Archives, in: Proceedings of the DLM-Forum on electronic records. European citizens and electronic information: the memory of the Information Society (= INSAR, Supplement 4), Luxembourg 2000, S. 272-278.

⁹ Die offizielle EAD-homepage wird von der Congress Library in Washington betreut: <<http://www.loc.gov/ead/ead.html>>. Die wichtigsten Handbücher sind die Tag Library und die Application Guidelines: Encoded Archival Description Tag Library, Version 1.0 (EAD Technical Documents No. 2), Chicago 1998

Werkzeuge der unterschiedlichsten Art können über das Internet heruntergeladen werden.¹⁰ Dabei hat die Tatsache, dass viele Hilfsprogramme lizenzfrei zur Verfügung stehen, unter anderem zur Folge, dass die Abhängigkeit von kommerziellen Softwareanbietern vergleichsweise gering ist. Mit dem Umstieg auf XML dürfte sie sogar noch weiter abnehmen.¹¹ Außerdem existiert inzwischen ein weit verzweigtes Anwender-Netzwerk, das einen intensiven Informationsaustausch ermöglicht, von dem vor allem Neueinsteiger profitieren können.¹²

Ein nicht geringer Teil des Erfolges der EAD dürfte auch auf die Flexibilität in der Anwendung zurückzuführen sein. Zwar hat die Erstellung der Findbücher bestimmten Regeln zu folgen, doch zeigt sich in der Praxis, dass in diesem Rahmen die Festlegung individueller, an den Erfordernissen der jeweiligen Institution ausgerichteter Praktiken und Regelwerke möglich bleibt. Große Freiheiten bestehen zudem im Bereich der Präsentation, da die EAD hierfür keine Richtlinien vorgibt. Schon jetzt ist bei

(<<http://lcweb.loc.gov/ead/tglib/tlhome.html>>); Encoded Archival Description Application Guidelines, Version 1.0 (EAD Technical Documents No. 3), Chicago 1999 (<<http://www.loc.gov/ead/ag/aghome.html>>).

¹⁰ Einen Überblick über Softwareprodukte für die Erzeugung, Präsentation, Indizierung von sowie die Recherche in EAD-Dokumenten bietet etwa <<http://jefferson.village.virginia.edu/ead/products.html>>.

¹¹ Um ein im SGML-Format vorliegendes EAD-Dokument öffnen zu können, werden spezielle Browser benötigt, die von kommerziellen Anbietern vertrieben werden (Panorama, Dynaweb). Ob nach dem Umstieg auf XML für die dann benötigten Stylesheets erneut kostenpflichtige Produkte auf den Markt gebracht werden und diese sich auch durchsetzen, bleibt abzuwarten.

¹² Siehe hierzu etwa die EAD-Help Pages: <<http://jefferson.village.virginia.edu/ead/sitesann.html>> oder die Übersicht auf der Homepage der Congress-Library: <<http://loc.gov/ead/eadsites.html>>. Gedruckte Erfahrungsberichte bieten folgende Sammelbände: Jackie M. Dooley (Hrsg.), Encoded Archival Description: Context, Theory, and Case Studies, Chicago 1998, hier S. 91-173; Archives and Museum Informatics, Volume 12 (1998), S. 169-305.

den Internet-Auftritten der beteiligten Institutionen eine große Vielfalt unterschiedlicher Oberflächen und Zugangsweisen festzustellen, ohne dass dadurch die gemeinsame Logik gänzlich verloren ginge. Mit Durchsetzung der XML-Version werden sich die bestehenden Freiheiten sogar noch erweitern, da die Darstellung auf dem Bildschirm dann über so genannte Style-Sheets gesteuert wird, die jeder Anbieter selbst definieren kann.

Aus archivischer Sicht ist vor allem zu begrüßen, dass EAD eine strukturierte Präsentation von Archivgut erlaubt und eine navigierende Recherche entlang der bestandsimmanenten Strukturen ermöglicht. Die Fähigkeit, Hierarchien wiederzugeben, macht den großen Fortschritt aus, den die EAD gegenüber ihrem Vorläufer, dem primär auf die Bibliothekswelt zugeschnittenen MARC (AMC)-Format, darstellt.¹³ Ein beachtlicher Vorteil ist es auch, dass sich die Online-Recherche nicht auf einzelne Findmittel, ja nicht einmal auf einzelne Archive beschränken muss. Die Standardisierung ermöglicht vielmehr die Schaffung zentraler Eingangsportale, die in Form gemeinsamer Datenbanken bzw. einer gemeinsamen Indexierung eine Bestände- und Institutionen übergreifende Recherche gewährleisten.¹⁴

III.

Für die Archive in Deutschland gibt es somit eine Reihe guter Gründe, sich des Näheren mit den Verwendungsmöglichkeiten der Encoded Ar-

¹³ Daniel Pitti, The Berkeley Finding Aid Project. Standards in Navigation: <<http://sunsite.berkeley.edu/FindingAids/EAD/arlpap.html>>.

¹⁴ Bedeutendstes Internet-Portal für archivische Findmittel ist die Datenbank der Research Library Group (RLG): <<http://www.rlg.org>>. Zentrale Zugänge eröffnen aber auch Online-Archive wie etwa das Online Archive of California: <<http://sunsite2.berkeley.edu/oac/>>.

chival Description zu beschäftigen. Andererseits ist nicht zu übersehen, dass einem konkreten Einsatz doch auch Hindernisse im Weg stehen. Diese sind nur zum Teil technischer Art. Vielmehr wurden - wie bereits angedeutet - in den Diskussionen mit den amerikanischen Teilnehmern auch spürbare Unterschiede in der archivischen Terminologie, Methodik und Praxis deutlich, die bei einer Evaluierung der EAD unbedingt berücksichtigt werden müssen.

So machte etwa der konkrete Versuch, ein mit Hilfe des in einigen Bundesländern verwendeten Softwaresystems MIDOSA-Online erstelltes Findmittel nach EAD zu konvertieren und auf dem Server der Research Library Group zu hinterlegen, deutlich, dass vor einer Adaptierung zahlreiche Detailfragen zu klären sind. Bei dem Bemühen um die Herstellung einer Konkordanz zwischen den Feldern der MIDOSA-Datenbank und den Markierungen der EAD waren einzelne Anforderungen nur schwer, manche vorerst noch nicht zu realisieren. Zwar zeigten sich auch Möglichkeiten, das System zu überlisten. Man wird sich aber fragen müssen, ob dies sinnvoll ist oder ob es letzten Endes nur dazu führt, falsche Vorstellungen zu provozieren.

Woraus ergeben sich aber die Schwierigkeiten bei der Übermittlung digitaler Erschließungsinformationen? Beschäftigt man sich eingehender mit dieser Frage, wird rasch deutlich, dass dabei keineswegs nur terminologische Hürden zu überwinden sind. So hat ein Teil der konstatierten Probleme auch mit der Tatsache zu tun, dass die Beständeerschließung in den Vereinigten Staaten überwiegend unter Zuhilfenahme von Textverarbeitungssystemen erfolgt, während im deutschen Sprachraum Datenbanklösungen vorherrschen. Zu berücksichtigen sind ferner die spezifischen Erschließungspraktiken der US-amerikanischen Archive, die ebenfalls nicht ohne Auswirkungen auf die Grundkonzeption der EAD geblieben sind. Abweichungen ergeben sich aber auch aus der für die

Vereinigten Staaten charakteristischen engen Zusammenarbeit von Archiven, Bibliotheken und Museen. An einigen besonders markanten Punkten soll dies im Folgenden näher erläutert werden.

Um mit den terminologischen Barrieren anzufangen, sei beispielhaft auf zwei zentrale Begriffe eingegangen, den Begriff "collection" sowie den Begriff "series". "Collection" würde man im Deutschen wohl am ehesten mit Sammlung übersetzen. Im Sprachgebrauch der amerikanischen Archivare beschreibt er allerdings nicht nur Sammlungsgut im engeren Sinne, also Flugblätter, Plakate, Fotos und museale Gegenstände, sondern auch Nachlässe und Verwaltungsschriftgut. Spricht man hingegen von Sammlungsgut in der dem deutschen Archivar geläufigen Bedeutung, wird in der amerikanischen Fachliteratur als Hilfsbezeichnung der Ausdruck "artificial collection" verwendet.¹⁵

Auch der Begriff "series" wird von den US-amerikanischen Archivaren in einem sehr spezifischen Sinne gebraucht. Mit "series" werden Aktenkomplexe bezeichnet, die als geschlossene Bestandsgruppen in die Archive übernommen wurden. Daher spiegeln sie nicht selten bestimmte Funktionen oder organisatorische Strukturen der abgebenden Stellen wider. So kann sich etwa die Überlieferung eines Krankenhauses aus zwei "series" zusammensetzen, aus den Patientenakten auf der einen und den Verwaltungsakten auf der anderen Seite. Wie in diesem Fall spielen bei der Entstehung solcher Aktenkomplexe oft formale Gründe eine wichtige Rolle. Ein weiteres Beispiel hierfür wären die in den Beständen amerikanischer Archive und Bibliotheken häufig anzutreffenden Protokoll- und Korrespondentenakten, die bei entsprechend langer Laufzeit als eigene Aktenserien in Erscheinung treten. Als Serien werden auch Teilbestände

¹⁵ Richard Rinehart, Anne Gilliland-Swetland, Editorial, in: *Archives and Museum Informatics*, Vol. 12 (1998), S. 171-176, hier S. 173f.

bezeichnet, die ihre Existenz archivalientypologischen Voraussetzungen oder rein lagerungstechnischen Gründen zu verdanken haben wie etwa die Bild- oder Kartenbestände.

Für die Adaption der EAD ist eine aus deutscher Sicht missverständliche Begriffsverwendung schon für sich gesehen nicht ganz unproblematisch. Zum einen existiert bislang keine deutsche Version und die wichtigsten Hilfsmittel liegen nur in der Originalsprache vor. Zum anderen können sich Unklarheiten und Fehlannahmen ergeben, wenn der Online-Zugang durch Länder und Institutionen übergreifende Portale und Datenbanken eröffnet werden soll, um so mehr, wenn die englische Version dabei als die kanonische Form für Kommunikation und Datenaustausch angesehen wird.¹⁶

Bei genauerem Zusehen wird ferner deutlich, dass das unterschiedliche Begriffsverständnis mit Abweichungen in der Erschließungspraxis korrespondiert, die ihrerseits einen beachtlichen Einfluss auf die Ausgestaltung der EAD gewonnen haben. Für ein vertieftes Verständnis derselben sind daher einige Grundkenntnisse über die in den USA gebräuchlichen Erschließungsmethoden erforderlich. Dabei fällt besonders ins Auge, dass die Ordnung und Verzeichnung des Archivguts nicht primär auf die einzelne archivalische Einheit abgestellt ist, sondern häufig kumulativ erfolgt. Das heißt, dass die Verzeichnung in aller Regel von oben nach unten erfolgt und auf eine Detailerschließung nicht selten verzichtet wird. Der Bestand wird zunächst als solcher beschrieben. Erst im Anschluss daran werden auch seine Haupt- und Untergruppen und - wenn überhaupt - die einzelnen Archivalieneinheiten erfasst. Ein weiteres Merkmal US-amerikanischer Erschließungspraxis ist es, die inneren Strukturen des

¹⁶ Daniel V. Pitti, Encoded Archival Description. An Introduction and Overview, in: D-Lib Magazine, Vol. 5, Number 11, November 1999 = <<http://www.dlib.org/dlib/november99/11pitti.html>> .

Verwaltungsschriftguts auch nach seiner Übernahme in das Archiv weitgehend unangetastet zu lassen, es also in der Form zu erschließen, in der es erwachsen ist. Schließlich fällt auf, dass zwischen der Erschließung von Nachlass- und Sammlungsgut und der von Verwaltungsunterlagen kein nennenswerter Unterschied gemacht wird.

Dieses Vorgehen findet seinen Niederschlag nicht nur in den Empfehlungen, die für das praktische Verfahren zur Erzeugung von EAD-Findmitteln gemacht werden, sondern hat auch Auswirkungen auf die Strukturierung der Verzeichnungsergebnisse. Soweit Bestände überhaupt hierarchisch gegliedert werden und man sich nicht mit einer linearen Auflistung begnügt, geschieht dies nicht nach den im deutschen Sprachraum üblichen Klassifikationsschemata, die die Zuständigkeiten und Aufgabengebiete einer Behörde differenziert widerspiegeln, sondern vielfach nach äußeren, das heißt nach formalen oder sogar nach rein physischen Kriterien. So orientiert sich die Bestandsstruktur - wie bereits angedeutet - meist an der von den Abgabestellen übernommenen Ordnung, auf der obersten Ebene also an den in den Behördenregistraturen gebildeten "series". Auch bei der Tiefengliederung spielen äußerliche Momente eine gewichtige Rolle. Dies geht nicht selten so weit, dass die Ablage- und Aufbewahrungsbehältnisse selbst als Gliederungselemente dienen. Komponenten der Strukturierung sind dann die Schachteln (container, box), die Aktendeckel oder -hefter (folder) oder auch die Filmrollen (reel). Dementsprechend beschreiben auch die der EAD zur Auszeichnung der Bestandsstrukturen verwendeten Markierungen (Tags) eher physische oder formale denn logische Einheiten oder Merkmale.¹⁷

¹⁷ Zur Strukturierung dienen unter EAD die so genannten "components" (= "information about the content, context and extent of a subordinate body of material"). Zur Charakterisierung dieser Tags siehe Encoded Archival Description Tag Library, Version 1.0 (EAD Technical Documents No. 2), Chicago 1998, S. 9f. und 70-86.

Die Bevorzugung einer eher summarischen Verzeichnung hat zudem zur Folge, dass auch der Umfang der erfassten Merkmale beziehungsweise deren Ausgestaltung nicht ganz den in Deutschland üblichen Erschließungsstandards entsprechen. So fehlen beispielsweise spezielle Felder für den Nachweis der Aktenzeichen und der Altsignaturen. Bei der Wiedergabe von Jahreszahlen lässt sich nur ein Teil der bei den deutschen Archiven gebräuchlichen Darstellungsformen realisieren. Umsetzungsschwierigkeiten traten auch bei der Beschreibung von Bandserien zutage. Eine genauere Aufschlüsselung der Akteninhalte in Form von Enthält- und Darin-Vermerken ist im Grunde eben sowenig vorgesehen wie die Einbindung von Zusatzinformationen zum Kontext, zur Bewertung oder zur Bearbeitung im Archiv.

Als weiteres Desiderat der EAD ist die fehlende Einbindung von Bestandsübersichten anzusehen. Vor allem bei größeren archivischen Einrichtungen, die die Überlieferung einer Vielzahl von Abgabestellen verwahren, stellen systematisch strukturierte und unmittelbar mit den Findmitteln verknüpfte Bestandsverzeichnisse eine erhebliche Zugangserleichterung dar. Noch größere Bedeutung erlangen sie, wenn die Benützung über zentrale Index- und Nachweisdatenbanken erfolgen soll. Zwar sind Bestandsübersichten auch in den USA gebräuchlich, doch begegnen diese vorwiegend noch in gedruckter Form. Soweit Online-Lösungen bereits existieren, wurden sie als proprietäre Anwendungen erstellt, die keine Verknüpfung mit der EAD erlauben.

Die Tatsache, dass in den USA vorwiegend Textverarbeitungssysteme zur Erschließung verwendet werden, hat zur Folge, dass die bereits existierenden EAD-Handbücher und -Hilfsmittel nahezu ausschließlich die Umwandlung von Textdateien in eine internetkompatible Form unterstützen. Liegen die digitalen Erschließungsinformationen hingegen in Form einer Datenbank vor, stellt das im Rahmen der EAD vorgesehene Verfah-

ren einen Umweg dar. Dieser ließe sich allerdings dann umgehen, wenn es gelänge, Schnittstellen zu definieren, die eine rasche und zuverlässige Generierung von EAD-Findmitteln aus der jeweiligen Datenbank ermöglichen. Da der Markt für archivische Erschließungssoftware in Deutschland von proprietären Lösungen beherrscht wird und auch bei der Erschließungspraxis regionale Eigenarten zu berücksichtigen sind, sind Schnittstellen dieser Art allerdings nur schwer standardisierbar. Vielmehr müssten sie für jedes Produkt neu entwickelt werden.

Auch die Online-Präsentation selbst lässt aus deutscher Sicht noch Wünsche offen. Angesichts der Fülle unterschiedlicher Angebote sind hier pauschale Aussagen aber schwierig. Gerade bei der Navigation scheinen aber Verbesserungen möglich zu sein. So bedienen sich zwar die meisten Online-Findbücher, die mehr bieten als eine rein lineare Auflistung, bereits der Rahmen- und Fenstertechnik, um die Rechercheergebnisse in einen größeren strukturellen Zusammenhang zu stellen. In aller Regel wird dabei aber allenfalls die nächsthöhere Verzeichnungebene angezeigt. Bei stärker untergliederten Beständen reicht dies aber nicht aus, um die jeweilige Position innerhalb des Findbuchs ausreichend deutlich zu machen. Wünschenswert wäre daher ein weiterer Rahmen, der es erlaubt, die Fundstelle in der Gesamtsystematik des Findbuchs zu verorten.

IV.

Beim Abschlusstreffen mit der amerikanischen Arbeitsgruppe wurde Einigkeit darüber erzielt, dass eine einfache Übernahme der Encoded Archival Description durch die deutschen Archive wegen der deutlich voneinander abweichenden archivischen Standards und wegen der anders gelagerten technischen Voraussetzungen zum jetzigen Zeitpunkt nicht ohne weiteres empfohlen werden kann. Als eines der größten Hemmnis-

se erwiesen sich dabei die Unterschiede in der Erschließungspraxis. Dazu kommt, dass es für die deutschen Archive, die ihre Verzeichnungsarbeiten weniger mit Textverarbeitungs- als mit Datenbank-Systemen bewältigen, inzwischen direktere Wege zur Präsentation von Online-Findmitteln gibt als den auf der Grundlage von EAD.

Vorbehalte sind auch gegenüber dem Vorschlag angezeigt, die EAD als universelles Standard-Austauschformat für Erschließungsinformationen zu nutzen. Für diesen Zweck wurde die Document Type Definition nicht entwickelt. Ihr Funktionsumfang ist daher zu klein, um wirklich alle Anforderungen zu erfüllen, die in einem solchen Fall zu stellen wären. Sollte dieses Ziel angestrebt werden, so wäre die Voraussetzung dafür eine umfassende Überarbeitung beziehungsweise Erweiterung der EAD, bei der die spezifischen Bedürfnisse aller daran interessierten Länder Berücksichtigung finden müssten.

Dies schließt allerdings keineswegs aus, die EAD als Grundlage für eine internationale Zusammenarbeit zu nutzen. Speziell bei Projekten, die die Einbindung von Findmitteln deutscher Herkunft in US-amerikanische Verbünde und Nachweissysteme zum Ziel haben, bieten sich interessante Perspektiven, da man in einem solchen Fall Einschränkungen bei der Funktionalität oftmals tolerieren können. Um sich diese Option offenzuhalten, würde es genügen, die im deutschen Sprachraum gebräuchlichen Datenbanksysteme um Exportschnittstellen zu erweitern, die eine Ausgabe der Erschließungsinformationen nach den Regeln der EAD erlauben. Damit könnte dem Wunsch nach einer Intensivierung der internationalen Zusammenarbeit mit vergleichsweise einfachen Mitteln Rechnung getragen werden. Für solche Gemeinschaftsprojekte böte sich nicht nur die Erstellung gemeinsamer Findmittel für grenzüberschreitende Überlieferungen an, sondern gerade auch der zentrale Nachweis für die

provenienzmäßig ungebundenen und daher schwer zu ermittelnden Nachlass- und Sammlungsbestände.

Unabhängig davon kann die Encoded Archival Description aber auch als lehrreiches Beispiel für die deutschen Archive dienen, die sich die Entwicklung von Online-Findmitteln zum Ziel gesetzt haben. So machen der breite Einsatz der EAD in den USA sowie das wachsende Interesse daran in einer Reihe weiterer Länder deutlich, welchen Stellenwert die Online-Präsentation von Erschließungsinformationen auf internationaler Ebene heute bereits besitzt. Auf diesem Gebiet ist für die Archive in Deutschland zweifellos ein Nachholbedarf zu konstatieren. Wie dieser zu bewältigen ist, dafür bietet die EAD ein gutes Beispiel.

Als Vorteile lassen sich auch die Flexibilität, vor allem aber die Interoperabilität der EAD ansprechen. So werden dem einzelnen Anwender ohne Zweifel genügend Spielräume für eine den eigenen Anforderungen entsprechende Ausgestaltung seiner Online-Findmittel gelassen. Weder der Umfang der zu erhebenden Erschließungsinformationen noch die Form der Präsentation sind fest vorgegeben. Mit dem Umstieg auf XML wird es sogar noch einfacher möglich sein, die Benutzeroberfläche individuell zu gestalten.

Am Beispiel der EAD kann aber zugleich gezeigt werden, welchen Zugewinn zentrale Nachweissysteme bieten können, die eine Bestände und Institutionen übergreifende Recherche ermöglichen. Diese wird durch eine fest vorgegebene Definition der für die Auszeichnung der Findbücher zur Verfügung stehenden Markierungen (Tags) gewährleistet. Von Bedeutung ist in diesem Zusammenhang aber auch die überregionale, inzwischen sogar internationale Zusammenarbeit der beteiligten Einrichtungen. Diese erschöpft sich nicht in der Tätigkeit der EAD-Working-Group, die für die Pflege der DTD zuständig ist, sondern setzt sich fort in der Schaffung gemeinsamer Eingangsportale und im Erfahrungsaus-

tausch der Anwender, der die Bereitstellung selbst entwickelter Hilfsmittel und Software-Werkzeuge mit einschließt. Das dabei entstandene Netzwerk kann Vorbildcharakter beanspruchen, auch wenn es in der Bundesrepublik wegen der hier bestehenden heterogenen Systemvoraussetzungen nicht ohne weiteres kopiert werden kann.

Alles in allem wird sorgfältig zu beobachten sein, wie sich die Encoded Archival Description weiterentwickelt. Wenn der derzeitige Eindruck nicht trügt, wird sich ihre zunehmende Verbreitung auf internationaler Ebene weiter fortsetzen. Die EAD-Working-Group umfasst inzwischen auch Mitglieder aus Großbritannien, Australien, Kanada und Frankreich. Eine Überarbeitung der EAD-Version 1.0 ist in Vorbereitung. Ihr Ziel ist es, Zusatzanforderungen zu realisieren, die gerade auch im Zuge der Internationalisierung zu Tage getreten sind. Es zeichnet sich ab, dass damit eine weitere Flexibilisierung der EAD einhergehen wird. Neu sind auch die Bemühungen, parallel dazu ein Kodierungsschema für Angaben über die Registraturbildner (Institutionen, Personen, Familien) zu entwickeln, das unter dem Begriff Encoded Archival Context (EAC) vorgestellt wurde und von Anfang an auf einer breiten internationalen Zusammenarbeit beruht.¹⁸

¹⁸ Siehe <<http://www.library.yale.edu/eac>>.

Gerald Maier und Thomas Fricke: Digitalisierung von Archivgut als Teil eines archivischen Online-Angebots - Workflow und Werkzeuge zur Digitalisierung von Archivgut. Ein Werkstattbericht¹

In unserer modernen Informationsgesellschaft bekommt das Schlagwort "Digitalisierung" in Verbindung mit dem Kommunikationsmedium Internet einen immer bedeutenderen Stellenwert für die Wissensvermittlung.

Als Reaktion auf die neuen Kommunikationsmedien beschäftigen sich auch Archive und Bibliotheken neben der Reprographie in Form der Mikroverfilmung zunehmend mit den Möglichkeiten der retrospektiven Digitalisierung von Archiv- und Bibliotheksgut. So besteht die Möglichkeit, interessierten Bürgern, Wissenschaftlern, Studenten und Schülern den Zugang zu den Quellen über Online-Medien wie dem Internet, aber auch offline in den Lesesälen der Archive und Bibliotheken oder über CD-ROM am heimischen PC zu ermöglichen. Dadurch können einerseits die wertvollen Originale und Unikate geschützt werden, andererseits erhält der Nutzer völlig neue Zugriffsmöglichkeiten, unabhängig von Ort und Zeit, auf Archivgut.

Ein archivisches Online-Angebot lässt sich bekanntlich in mehrere Ebenen gliedern:

1. Archivübergreifende Portale mit Suchmaschinen,
2. Archiv-Homepage mit Grundinformationen,
3. Online-Beständeübersicht, evtl. mit sachthematischen Bezügen,
4. Online-Findbücher und
5. digitalisiertes Archivgut.

¹ Stand: 01.10.2001

Mit einem archivischen Online-Angebot wird ein sog. Mehrwert gegenüber einer konventionellen Präsentation erreicht durch eine erweiterte Zugänglichkeit unabhängig von Ort und Zeit, durch die Integration der Dokumente in Tektonik und Kontext (u.a. Hintergrundinformationen, didaktische Aufbereitung, Transkriptionen) und durch einen mehrdimensionalen Zugriff mit erweiterten Navigationsmöglichkeiten (u.a. Strukturansicht, Volltextsuche, Indices, Querverweise).

Dieser Beitrag beschäftigt sich konkret mit den beiden untersten Stufen, d.h. den Online-Findbüchern mit digitalisierten Archivalien, innerhalb eines archivischen Online-Angebots. Im Rahmen eines von der Deutschen Forschungsgemeinschaft geförderten Projekts der Landesarchivdirektion Baden-Württemberg werden hierfür neue Werkzeuge und Möglichkeiten der Präsentation erarbeitet.

Projekt „Workflow und Werkzeuge zur digitalen Bereitstellung größerer Mengen von Archivgut“

Voraussetzung für die rationelle Digitalisierung von Archivgut sind weitgehend automatisierbare Arbeitsabläufe zur Herstellung optimaler digitaler Master und zur Einbindung der digitalen Formen in Online-Findbücher.

Im Rahmen des DFG-Projekts „Workflow und Werkzeuge zur digitalen Bereitstellung größerer Mengen von Archivgut“² versucht die Landesarchivdirektion Baden-Württemberg Lösungen für die in Online-Findbüchern integrierte Präsentation digitalisierter Archivalien zu entwickeln. Erfahrungen mit der Digitalisierung von Archivgut konnte die Landesarchivdirektion bereits im Vorgängerprojekt „Digitale Konversionsformen“ erwerben. Vorrangiges Projektziel war dabei die Entwicklung von Methoden und technischen Verfahren für die Digitalisierung von Ar-

² URL: <http://www.lad-bw.de/workflow>

chiv- und Bibliotheksgut mit dem Schwerpunkt der Digitalisierung vom Mikrofilm³.

Die dabei entwickelten Präsentationsmodelle und Testpräsentationen sind im Internet-Angebot der Landesarchivdirektion Baden-Württemberg zu finden⁴.

Der Workflow für die Digitalisierung von Archivgut



Abb. 1: Workflow für die Digitalisierung von Archivgut

³ Die Ergebnisse dieses Projekts sind ausführlich dokumentiert in Hartmut Weber/Gerald Maier (Hgg.), Digitale Archive und Bibliotheken. Neue Zugangsmöglichkeiten und Nutzungsqualitäten, Stuttgart 2000 (= Werkhefte der Staatlichen Archivverwaltung Baden-Württemberg, Serie A Landesarchivdirektion, Heft 15).

⁴ URL: <http://www.lad-bw.de/digpro>.

Um eine rationelle Mengenverarbeitung und eine Integration der digitalen Formen in Online-Findmittel zu erreichen, werden im Projekt „Workflow“ zum einen weitgehend automatisierbare Arbeitsabläufe zur Herstellung optimaler digitaler Master über das Medium Mikrofilm, zum anderen Autorensystem gestützte Präsentationsmodule für die objektgerechte Präsentation von Urkunden mit Siegeln, für Akten und für Zeitungen entwickelt.

Der im Rahmen des Projektes entwickelte Digitalisierungs-Workflow besteht aus folgenden Prozessschritten:

1. Erschließung und Auswahl der Objekte/Dokumente⁵, die für eine Online-Präsentation vorgesehen sind, mit einer datenbankbasierten Erschließungssoftware
2. Verfilmung der Objekte/Dokumente unter Berücksichtigung der objektgerechten Bildqualität
3. Digitalisierung der mikroverfilmten Objekte/Dokumente unter Berücksichtigung der objektgerechten Bildqualität
4. Bildorganisation unter Verwendung von Metadaten
5. Kontrolle der Bildqualität
6. Automatisierte Aufbereitung der digitalisierten Masterformen für eine Präsentation
7. Präsentation digitalisierter Objekte im Kontext eines archivischen Online-Angebots durch Einbindung der bearbeiteten digitalen Nutzungsformen in Online-Präsentationsmodule

⁵ Objekt = eine Titelaufnahme im Findbuch; Dokument = einzelne Seite/einzelnes Bild eines Objekts.

Als Basis und zur Steuerung des Workflows dient ein datenbankbasiertes Autorensystem zur Bearbeitung und Verwaltung der Ausgangsmaterialien, d.h. sowohl der Erschließungsdaten als auch der als Bilddaten vorliegenden digitalisierten Objekte. Dabei werden bereits vorhandene Erschließungsdaten mit Metadaten von digitalisierten Objekten bzw. Dokumenten verbunden. Die Datenbank enthält u.a. folgende Informationen:

- a) Erschließungsdaten
- b) Metadaten für die Mikroformen und Verwaltung der digitalen Masterformen
- c) Metadaten für die Präsentation

Das innerhalb des Projekts entwickelte Autorensystem arbeitet in einer Client-Server-Umgebung unter Verwendung des Datenbanksystems MySQL, der Webservererweiterung PHP und dem Apache Webserver⁶. Es wird ergänzt durch Standardsoftware für die Bildbearbeitung.

⁶ Zu MySQL siehe <http://www.mysql.org> und <http://www.mysql.com>, zu PHP siehe <http://www.php.org> und zum Apache-Webserver siehe <http://www.apache.org>.

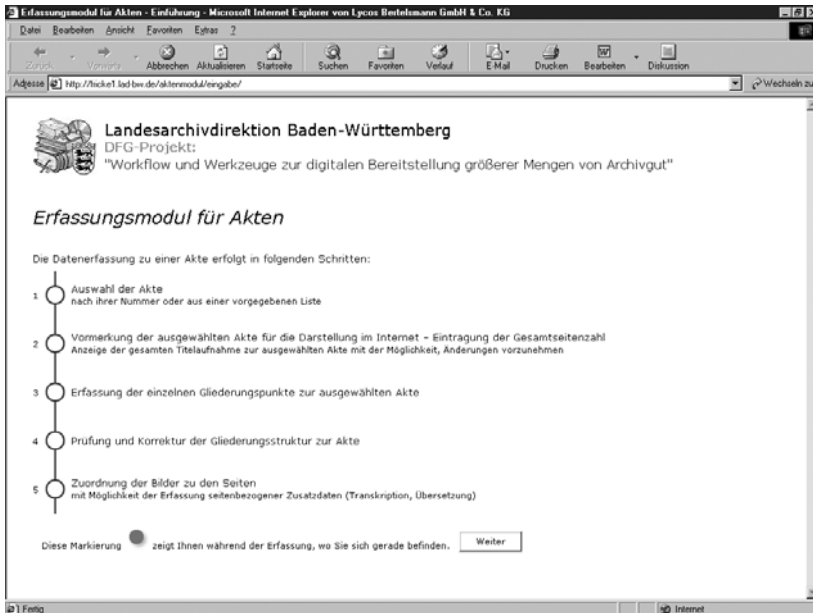


Abb. 2: Autorensystem

Die Visualisierung der aufbereiteten Daten erfolgt über archivalienspezifische Präsentationsmodule, die in Online-Findbücher integriert werden.

Erschließung und Auswahl der zu digitalisierenden Objekte/Dokumente

Abbildung 3 zeigt das Autorensystem für die Erschließung von Dokumenten. Oben ist ein Prozessdiagramm mit den Schritten 0 bis 5 dargestellt: 0 Ablauf, 1 Akte auswählen, 2 MIDOSA-Daten prüfen (Akte fürs Web vormerken, Seiten angeben), 3 Dokumentengliederung erfassen, 4 Dokumentengliederung prüfen und ordnen, 5 Seiten überprüfen (Bilder zuordnen). Darunter steht der Titel 'MIDOSA-Daten zur Akte: E 170 a 1'. Ein Kontrollfeld fragt, ob die Akte im Web präsentiert werden soll (Ja/Nein), wobei 'Nein' ausgewählt ist. Ein Feld für die Gesamtzahl der paginierten Seiten ist mit '3' besetzt. Der Hauptteil der Abbildung ist ein Formular mit verschiedenen Eingabefeldern: 'Akte typ' (SA), 'Archiv' (leer), 'Bestand' (E 170 a), 'Bearbeiter' (bu), 'Ebene' (F0), 'Druckcode' (a), 'Satz ID' (1), 'Ident.Nr' (2000 A F0), 'Bestellnummer' (BU), 'Titel' (Protokolle der Patentkommission), 'Enthalt/Darin' (leer), 'Laufzeit' (1848 - 1961), 'Umfang' (1 Bu), 'Klassifikation' (01), 'Endprovenienz' (leer), 'Vorprovenienz' (leer), 'Vorsignaturen' (1: , 2: , 3: , 4:), 'Angaben zur Bewertung' (leer), 'Angaben zur Erschließung' (leer), 'Angaben über Schäden' (leer), 'Hintergrundinformationen' (leer), 'Bemerkungen' (leer), 'Sortierfelder' (1. Begriff, 2. Begriff, 3. Begriff).

Abb. 3: Autorensystem. Erschließungsmodul

Die Erschließung erfolgt im Rahmen des Projekts über ein in das Autorensystem integriertes Erschließungsmodul, das auf einem erweiterten Datenmodell der Erschließungssoftware MIDOSA 95 bzw. MIDOSA für Windows basiert⁷. Die Dateneingabe erfolgt über einen Webbrowser.

⁷ Zu MIDOSA siehe <http://www.midosa.de>.

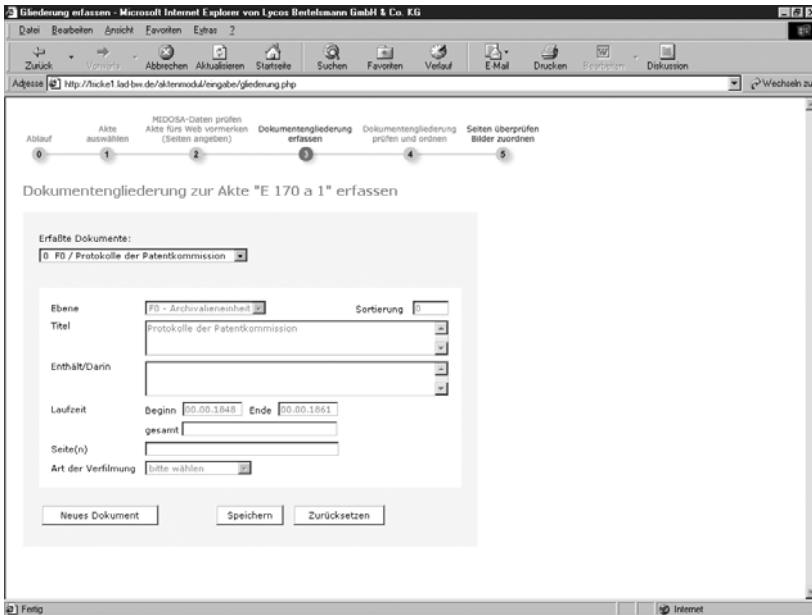


Abb. 4: Autorensystem. Erschließungsmodul

Das Erschließungsmodul ist mit einem weiteren, auf der selben Technik basierenden Modul für die Erfassung von Metadaten verbunden. Hiermit können während der Erschließung oder in einem zweiten Arbeitsgang diejenigen Archivalien oder Archivalieneinheiten ausgewählt werden, die für eine digitale Online-Präsentation vorgesehen sind.

Ablauf Akte auswählen MIDOSA-Daten prüfen
 Akte für Web vormerken (Seiten angeben) Dokumentengliederung erfassen Dokumentengliederung prüfen und ordnen Seiten überprüfen
 0 1 2 3 4 5

Dokumentengliederung zur Akte "E 170 a 1" ordnen und prüfen

► Seitenbelegung ◀

Änderungen speichern Zurücksetzen

Ebene	Sortierung	Titel	Enthält/Darin:	Laufzeit	Seiten	Art der Verfilmung	
F0		Protokolle der Patentkommission		1848 - 1861	6		Löschen
	G0	1 (2)	Patentschrift	01.01.1848 - 01.09.1848	1-2	35mm Halbton	Löschen
	H0	2 (5)	Zeichnung	00.00.1850 -	6	Makrofische Farbe	Löschen
	H0	3 (3)	Beilage	02.02.1861 -	4	35mm Halbton	Löschen
	J0	4 (4)	Notiz	01.01.1860 -	5	35mm Halbton	Löschen

Seitenbelegung

Seiten	Gliederungseinträge
1	1 - G0
2	1 - G0
3	
4	3 - H0
5	
6	2 - H0

Abb. 5: Autorensystem. Erschließungsmodul

Im Erschließungsmodul ist auch die schon im Datenmodell von MIDOSA 95 bzw. MIDOSA für Windows angelegte Möglichkeit der Stufenerschließung ausgebaut worden und kann für die gesonderte Beschreibung einzelner zu einer Titelaufnahme gehörender Vorgänge oder Dokumente verwendet werden, für die eine digitale Präsentation vorgesehen ist.

Verfilmung der Objekte

Der hier beschriebene Workflow berücksichtigt die Digitalisierung von Mikroformen. Die zu präsentierenden Originalobjekte wurden daher zunächst nach bestimmten Vorgaben unter Berücksichtigung der objektgerechten Bildqualität auf 35 mm Rollfilme bzw. Makrofiches mikroverfilmt. Dabei kommt auch der Farbmikrofilm zum Einsatz⁸. Die Verfilmung erfolgt teilweise in Eigenleistung durch das Institut für Erhaltung von Archiv- und Bibliotheksgut der Landesarchivdirektion Baden-Württemberg in Ludwigsburg, teilweise durch einen externen Dienstleister.

Digitalisierung der mikroverfilmten Objekte/Dokumente unter Berücksichtigung der objektgerechten Bildqualität

Das Scannen der mikroverfilmten Objekte erfolgt hauptsächlich durch externe Dienstleister. Zu den wichtigen Faktoren im Vorfeld des Scanvorgangs gehört die Auswahl der objektgerechten Scanauflösung und Farbtiefe, verbunden mit der Wahl des geeigneten Dateiformats, um jeweils eine optimale Bildqualität zu erreichen⁹.

Die Bildqualität von digitalen Formen lässt sich hauptsächlich an folgenden Kriterien festmachen:

1. an der Bildschärfe:

Ziel ist die Wiedergabe des kleinsten abzubildenden Teils eines Bildes, die Wiedergabe scharfer Konturen und Lesbarkeit von Schriften.

⁸ Zur Mikroverfilmung von Archivgut als Basis für eine Filmdigitalisierung siehe Peter Exner, Verfilmung und Digitalisierung von Archiv- und Bibliotheksgut, in: Maier/Weber, Digitale Archive und Bibliotheken, S. 113-127.

⁹ Siehe dazu Gerald Maier, Qualität, Bearbeitung und Präsentation digitaler Bilder, in: Weber/Maier, Digitale Archive und Bibliotheken, S.129-178.

2. an der Tonwiedergabe:

Ziel sind Helligkeitsabstufungen, die möglichst denen der Originalvorlage entsprechen (Einfluss der Farbtiefe).

3. an der Farbwiedergabe:

Ziel ist eine ausgewogene Farbwiedergabe, die derjenigen der Originalvorlage entspricht (Einfluss der Farbtiefe).

4. an mögliche Bildstörungen:

Ziel ist die Vermeidung ungewollter, zufälliger Bildstörungen (z.B. Rauschen, Moiré, Streupixel).

Die Bildschärfe ist das hauptsächliche Qualitätsmerkmal bei bitonalen Scans, also bei der Digitalisierung von maschinenschriftlichen bzw. gedruckten Textvorlagen und Strichzeichnungen. Sie ist vor allem von einer ausreichenden Scanauflösung abhängig, aber auch von einer korrekten Fokussierung beim Scannen.

Je nach Beschaffenheit der Ausgangsformen, d.h. der Filme und der darauf abgebildeten Originalobjekte, ist eine höhere oder weniger hohe Scanauflösung notwendig. Ein Hauptkriterium für die Höhe der Scanauflösung ist bei der Digitalisierung von Archivgut dabei die Lesbarkeit der Dokumente. Für die zunächst beim Digitalisieren zu erstellenden Masterformen, die später zu Präsentationsformen verarbeitet werden, sind 600 dpi für 1 Bit Scans und 300-400 dpi für 8-48 Bit Scans bezogen auf eine Ausgangsgröße von DIN A4 sinnvoll.

Neben der Bildschärfe ist die Tonwiedergabe ein zweites wichtiges Qualitätskriterium. Die korrekte Tonwiedergabe ist vor allem ein wichtiges Qualitätskriterium für Graustufen- und Farbscans. Bei 1 Bit-Scans mit den beiden Tonwertstufen schwarz und weiß ist dagegen der sogenannte Schwellenwert für eine korrekte Wiedergabe entscheidend.

Zunächst ist für eine korrekte Ton- und Farbwiedergabe die Wahl einer objektgerechten Farbtiefe entscheidend:

- Eine Farbtiefe von 1 Bit (bitonal) ist geeignet für AHU-Filme mit Druckschriften, Druckschriften im Original, Laserausdrucke und Strichzeichnungen.
- Eine Farbtiefe von 8/10/12 Bit (Graustufen) und 24/36/48 Bit (Farben) ist erforderlich für AHU-Filme mit Halbtonvorlagen, Halbton- und Farbfilmen von Originalobjekten mit Handschriften, Bleistiftzeichnungen, Farbstift, Gewebefarbband, Illustrationen oder Fotografien.

Die visuelle Qualitätskontrolle der von Dienstleistern erstellten digitalen Formen hat auch gezeigt, dass für eine korrekte Ton- und Farbwiedergabe die Einführung eines sogenannten Colormanagements in Verbindung mit einer Gerätecharakterisierung und -kalibrierung eine unabdingbare Voraussetzung ist. Dies gilt vor allem für die Digitalisierung von sequentiellen Dokumenten eines Objektes oder gleichartiger Objekte. Im Rahmen des Projekts „Workflow und Werkzeuge“ werden daher auch Lösungen für eine praktikables Farbmanagement bei der Filmdigitalisierung von Archivgut erarbeitet.

Einfluss auf die Qualität und Dateigröße digitaler Formen hat auch die Wahl eines geeigneten Dateiformats in Verbindung mit Komprimierungstechnologien zur Reduktion der Dateigröße.

Bei der Frage nach der Verwendung eines Dateiformats muss unterschieden werden zwischen dem sog. digitalen Master und digitalen Nutzungsformen. Die Anforderungen, die an den digitalen Master gestellt werden, sind aus der Art der Digitalisierungsvorlagen abzuleiten. Auf jeden Fall soll er als Archivierungsversion eine verlustfreie Speicherung ohne Qualitätsverluste ermöglichen.

Die digitalen Nutzungsformen werden durch Konvertierung und Bearbeitung des Masters erstellt. Sie genügen niedrigeren Qualitätsanforderungen, benötigen aber zugleich eine möglichst geringe Dateigröße, um den raschen Datentransfer in Online-Medien wie dem Internet zu gewährleisten.

Für digitale Farb- und Graustufen-Master (8-48 Bit) bietet sich das TIF-Format in seiner unkomprimierten Variante als momentaner Quasi-Standard an, während es als TIFF-G4 mit verlustfreier Kompression für 1 Bit-Master sinnvoll erscheint. Die Lage könnte sich aber mit einer weiteren Verbreitung des PNG-Dateiformats ändern, das ebenfalls aufgrund seiner verlustfreien Kompression und anderer Vorteile nicht nur ein zukünftiges Dateiformat für digitale Nutzungsformen werden könnte, sondern bei entsprechender Weiterentwicklung (z.B. Möglichkeit der Einbindung von Farbprofilen) auch für digitale Master.

Bei der Erstellung von digitalen Nutzungsformen für Präsentationen ist man momentan immer noch mit Verwendung des JPEG-Dateiformats auf der sicheren Seite, da dieses Format auch von älterer Browser-Software gelesen werden kann. Das GIF-Format (Graphics Interchange Format) – geeignet für 1 Bit-Bilder und für 8 Bit-Graustufen- und Farbbilder – wird aufgrund seines lizenzpflichtigen Komprimierungsalgorithmus nicht empfohlen. Die vorteilhafte Eigenschaft des GIF-Formats gegenüber dem JPEG-Format, nämlich die Bewahrung einer guten Konturschärfe bei Schriftvorlagen trotz Kompression, besitzt auch das universalere PNG-Dateiformat. Daneben gibt es die bisher noch proprietären Dateiformate mit Wavelet-Komprimierungsalgorithmen, die bei akzeptabler Qualität eine beachtliche Komprimierung erreichen.

Die Filmdigitalisierung von 35 mm-Mikrorollfilmen im bitonalen und Graustufen-Modus erfolgt mit speziellen Rollfilmscannern, bei Makrofilm auch durch Flachbettscanner mit Durchlichteinrichtung.



Abb. 6: Scanner für 35mm-Mikrorollfilm

Für die Digitalisierung von 35 mm Farbmikrofilmen und Farbmakrofiches wird bisher normalerweise ein vom Dienstleister umgebauter Kamerascanner mit Durchlichteinrichtung oder ein Flachbettscanner mit Durchlichteinrichtung verwendet.

Bildorganisation unter Verwendung von Metadaten

Als Ergebnis des Scannens entstehen einzelne Bilddaten. Diese müssen nach bestimmten Kriterien organisiert und verwaltet werden, damit in den folgenden Schritten die Bildqualität kontrolliert werden kann und die Bilder rationell weiterverarbeitet werden können.

Entscheidend für die Verwaltung digitaler Formen ist die Identifikation der einzelnen Objekte mit ihren Dokumenten. Wichtig – auch aus archivfachlicher Sicht – ist dabei, dass der Zusammenhang einzelner Dokumente eines Objekts erhalten bleibt u.a. durch die Wahl entsprechender Verzeichnisstrukturen und Dateinamen.

Die Identifikation einzelner Objekte und Seiten erfolgt über den Dateinamen der digitalen Formen. Dabei setzen sich die Dateinamen aus der mit einer Bildkennung verbunden Bestellsignatur zusammen.

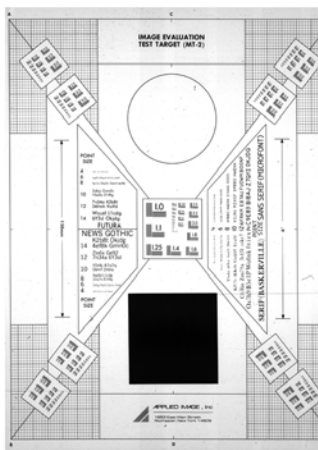
Kontrolle der Bildqualität

Im nächsten Schritt erfolgt die Qualitätskontrolle der gescannten Bilder. Kontrolliert werden müssen v.a. Bildschärfe, Tonwiedergabe, Farbwiedergabe und mögliche Bildstörungen.

Um eine gleichbleibende Bildqualität zu erreichen müssen schon im Vorfeld Maßnahmen zur Qualitätssicherung vorgenommen werden. Eine solche wird erreicht durch

- visuelle Qualitätskontrolle,
- die Berücksichtigung von Faktoren, die die Bildqualität beeinflussen
- und die Anwendung erprobter, normgerechter Bearbeitungsverfahren (z.B. richtige Reihenfolge verschiedener Bearbeitungsschritte in einem Workflow von der Mikroverfilmung über die Filmdigitalisierung bis hin zur Bildbearbeitung; Einsatz von Colormanagement)

Als Kontrollmethoden für die Bildschärfe dienen die Beurteilung von Testtafeln mit Testzeichen und Testbildern und das Testen von verschiedenen Schriftgrößen auf Lesbarkeit. Dabei wird auch überprüft, ob der vorgegebene QI (Quality Index)-Wert erreicht wird.



Testtafel „Image Evaluation Test Target (MT-2)“ mit ISO-Testzeichen Nr. 2

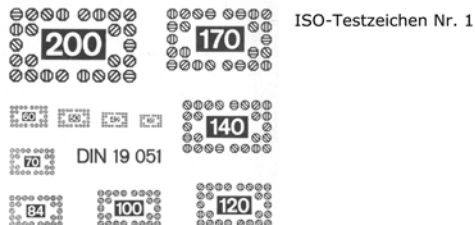
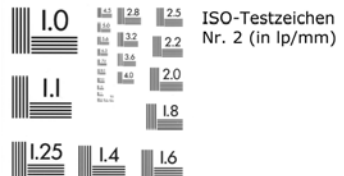


Abb. 7: Kontrolle der Bildschärfe über Testzeichen

Sinnvolle Kontrollmedien für die Beurteilung der Bildschärfe sind ein hochauflösender Monitor mit einer Farbtiefe von 24 Bit, in Verbindung mit einer Viewer-Software, die eine visuelle Qualitätskontrolle in verschiedenen Vergrößerungsstufen ermöglicht. Bei bitonalen Bildern erfolgt die Kontrolle zusätzlich über einen mit Laserdrucker erstellten Papierausdruck.

Die Tonwiedergabe - neben der Bildschärfe das zweite wichtige Qualitätskriterium - ist schwierig zu kontrollieren, da alle Komponenten eines Digitalisierungs-Systems Einfluss auf sie haben. Bei einem gescannten Bild muss im Vergleich zur Scanvorlage hinsichtlich der Tonwiedergabe folgendes geprüft werden:

- die gleichmäßige Detailabbildung in hellen und dunklen Stellen des Bildes,
- ein optimaler Gradationsverlauf,
- die Existenz signifikanter im Original enthaltener feiner Details.

Aufbereitung der digitalisierten Masterformen für eine Online-Präsentation

Nach der Qualitätskontrolle ist normalerweise der nächste Schritt innerhalb eines Digitalisierungs-Workflows das sogenannte Image Enhancement und die Aufbereitung des Bildmaterials für eine Online-Präsentation.

Wichtige Bereiche des Image Enhancements sind:

- Tonwertkorrektur
- Farbkorrektur
- Schärfenkorrektur

- Beseitigung von Artefakten (Bildstörungen wie Moiré, Störpixeln, Bildkratzern)
- Korrektur der Bildlage von schiefen Bildern
- Freistellen/Entfernen unerwünschter Ränder
- Trennung doppelseitiger Bilder
- Verkleinern von Bildern

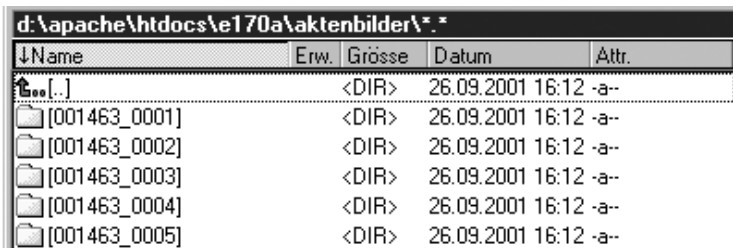
Im Rahmen des Projekts „Workflow und Werkzeuge“ werden diese Bereiche weitgehend automatisiert und im sog. Batch-Betrieb abgearbeitet. Daher werden die vom Dienstleister gelieferten digitalen Masterformen einer automatischen Bildaufbereitung unterzogen. Dabei kommen Standardsoftware aber auch eigens dafür programmierte Werkzeuge zum Einsatz. Wichtig ist dabei auch die Zuordnung der Bilder zu den entsprechenden Erschließungsinformationen. Dies geschieht mittels des schon genannten Autorensystems.

Einbindung der bearbeiteten digitalen Nutzungsformen in Online-Präsentationsmodule

Für die Präsentation digitalisierter Objekte im Kontext eines archivischen Online-Angebots werden im Rahmen des DFG-Projekt „Workflow und Werkzeuge“ Präsentationsmodule für Urkunden mit Siegeln, für Akten und für Zeitungen entwickelt. Dabei werden diese Module in die dafür angepasste und weiterentwickelte Software „MIDOSA online Findbuch“ integriert¹⁰.

¹⁰ MIDOSAonline ist aus dem Verzeichnungsprogramm MIDOSA 95 heraus entstanden. Teile der Entwicklung wurden von der Deutschen Forschungsgemeinschaft gefördert. Es wird von der Entwicklungspartnerschaft PARSIFAL aus Landesarchivdirektion Baden-Württemberg, Archivschule Marburg und Bundesarchiv gepflegt und weiterentwickelt.

Die für den Online-Zugriff aufbereiteten Bilddateien werden auf dem Server in einem Filesystem angeordnet. Datei- bzw. Ordernamen dienen als Kennung. Sie enthalten also die Signatur der Archivalieneinheit und die physikalische Seitenzahl. Die Namen werden später bei der Abfrage durch den Internet-Benutzer vom Präsentationsmodul analysiert und somit die Bilddaten mit den Datenbankeinträgen in Beziehung gebracht.



Name	Erw.	Grösse	Datum	Attr.
↑...[...]		<DIR>	26.09.2001 16:12	-a--
[001463_0001]		<DIR>	26.09.2001 16:12	-a--
[001463_0002]		<DIR>	26.09.2001 16:12	-a--
[001463_0003]		<DIR>	26.09.2001 16:12	-a--
[001463_0004]		<DIR>	26.09.2001 16:12	-a--
[001463_0005]		<DIR>	26.09.2001 16:12	-a--

Abb. 8: Struktur der Bilddaten im Filesystem

Wenn die Bilder bereits durch den Dienstleister, der die Verfilmung und Digitalisierung durchgeführt hat, mit entsprechenden Namen versehen wurden, können die Aufnahmen automatisch den Datensätzen der Datenbank zugeordnet werden. Sonst muss das Zuordnen bzw. Umbenennen der Bilder mithilfe des Erfassungsmoduls erfolgen.

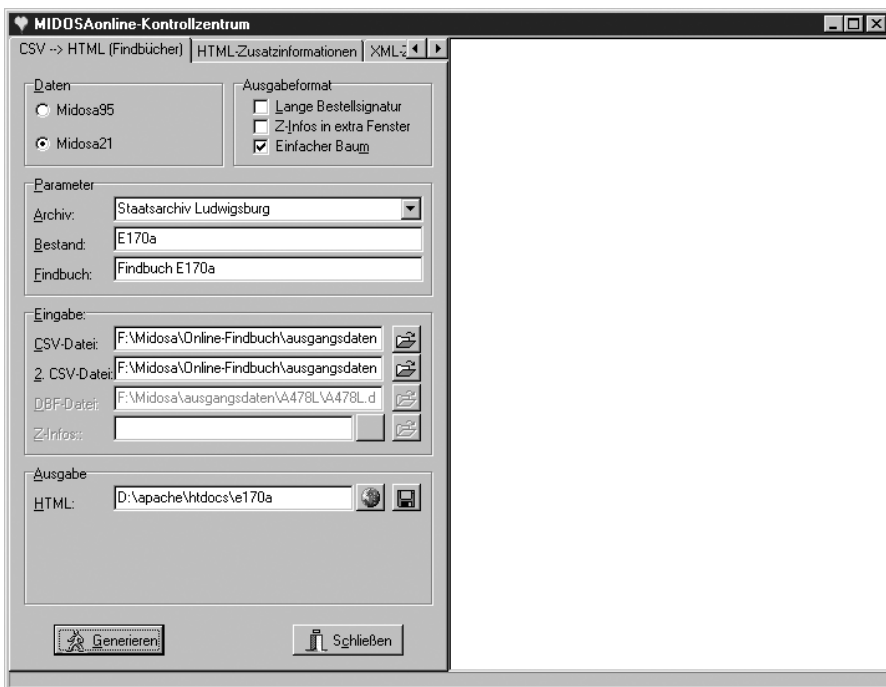


Abb. 9: MIDOSAonline Findbuch-Generator

Zum Schluss wird aus der Datenbank die aus HTML-Files bestehende, statische Strukturansicht mit dem Generator für MIDOSA online erzeugt. Für die Volltextsuche wird die Datenbank mit den Erschließungsinformationen zusätzlich zum Filesystem auf dem Server hinterlegt. Zu einem späteren Zeitpunkt wird die Generierung von statischen HTML-Files nicht mehr notwendig sein, da die Strukturansicht dann dynamisch aus der Datenbank generiert wird. Die Konfiguration des Präsentationsmoduls auf dem Webserver schließt sich daran an.

Die Software „MIDOSA online Findbuch“ bietet inklusive der speziellen Anpassungen durch die Landesarchivdirektion Baden-Württemberg folgende Möglichkeiten der Navigation und Recherche:

1. Kontextorientierte, strukturierte Navigation im HTML-Filesystem über Symbole und Hyperlinks
2. Recherche und Navigation über dynamisch generierte Gliederungsseite (Site Map)
3. Recherche über Index-Funktion
4. Recherche über datenbankgestütztes Volltextsuchmodul
5. Recherche über Internet-Suchmaschine für eine Suche im HTML-Filesystem außerhalb der Beständeübersicht im Rahmen einer Portal-lösung

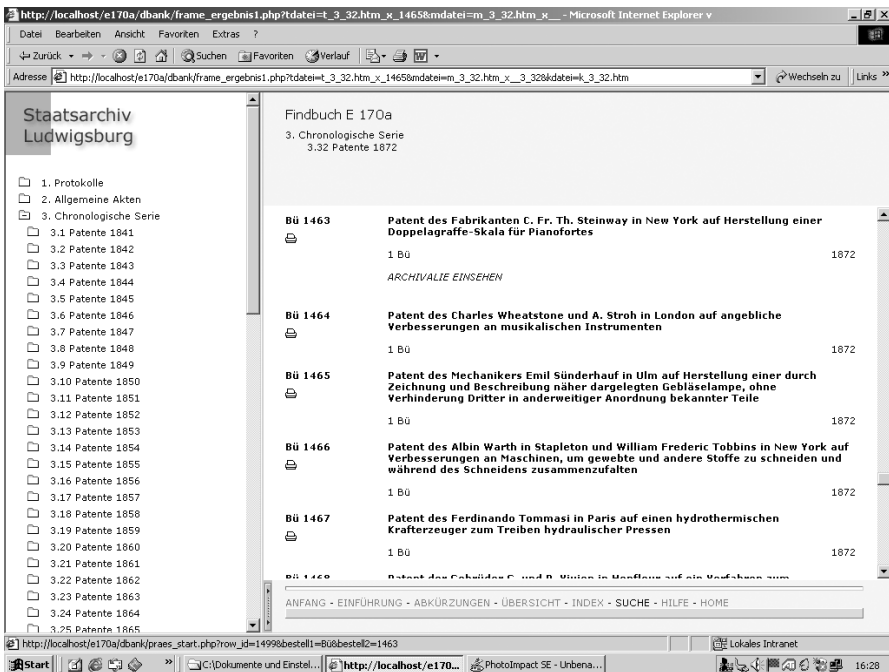


Abb. 10: MIDOSAonline Findbuch. Strukturansicht

Als Beispiel für die kontextorientierte Online-Präsentation von Akten wurde im Rahmen des Projekts der Bestand „Württembergische Patentkommission der Zentralstelle für Gewerbe und Handel 1848-1877 des Staatsarchivs Ludwigsburg gewählt¹¹.

Die einzelnen Patentakten dieses Bestands, die sich u.a. mit der Weiterentwicklung von Dampfmaschinen, der Textilverarbeitungstechnik, der Eisenbahntechnik oder mit der Öfenkonstruktion beschäftigen, enthalten vielfältige Beilagen u.a. in Form von Zeichnungen und Skizzen, teils einfarbig teils mehrfarbig.

Innerhalb des Projekts „Workflow und Werkzeuge“ werden in Auswahl Akten des Bestands nach dem oben beschriebenen Workflow digitalisiert und über ein angepasstes „MIDOSAonline Findbuch“ präsentiert.

Dem Findmittel liegt eine Datenbank zugrunde. Es vereint zwei verschiedene Möglichkeiten der Präsentation und Recherche:

1. Strukturansicht:

Für eine strukturorientierte Recherche erfolgt die Präsentation der Datenbank-Inhalte in Form eines statischen Dateisystems mit einer zweistufigen Klassifikation nach dem Jahr des Patentverfahrens und den Namen der Erfinder in alphabetischer Reihenfolge

2. Datenbanksuche:

Die Datenbanksuche erfolgt nach verschiedenen Kriterien und in Kombination verschiedener Suchbegriffe über Eingabeformulare. Die Suchergebnisse werden in einer festgelegten Form ausgegeben.

¹¹ Staatsarchiv Ludwigsburg, Bestand E170a.

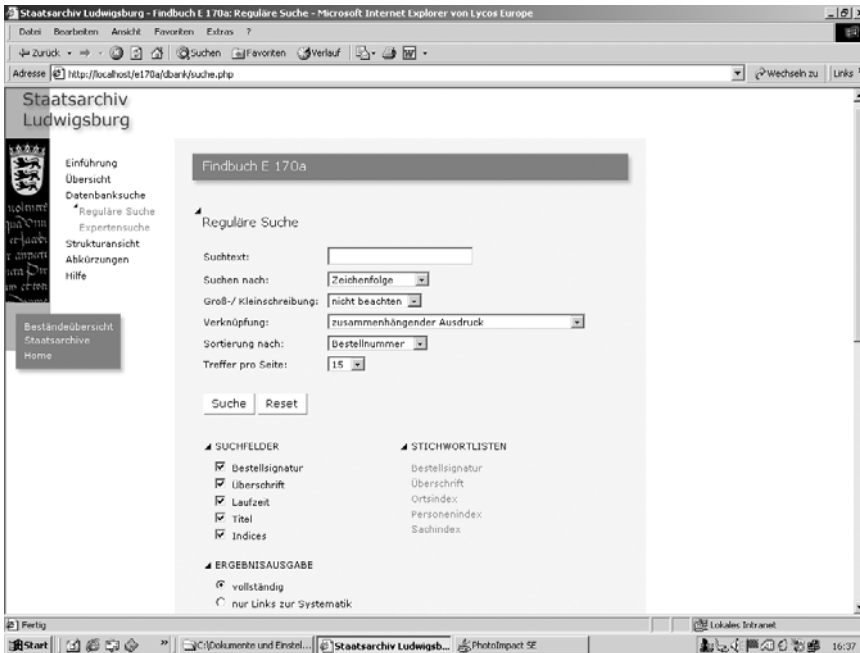


Abb. 11: MIDOSAonline Findbuch. Datenbanksuche

Der Zugriff auf die digitalisierten Objekte erfolgt innerhalb des Online-Findbuchs folgendermaßen: Über Hyperlinks unterhalb der Titelaufnahmen - dies gilt sowohl für die Strukturansicht als auch für den Datenbank-Report – können digitalisierte Objekte und Transkriptionen als Fak-similes in separaten Fenstern eingeblendet werden.

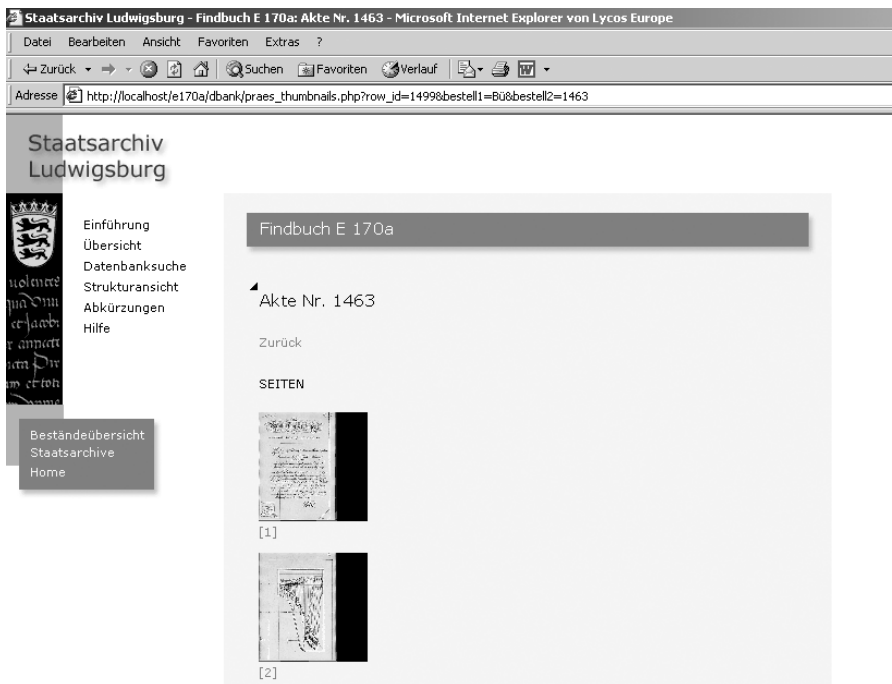


Abb. 12: MIDOSAonline Findbuch. Auswahl über automatisch generierte Vorschaubilder

Zunächst erhält der Nutzer ein Auswahlmenu, in dem einzelne Dokumente entweder durch ein Pulldown-Menu oder durch automatisch generierte Vorschaubilder ausgewählt werden können. Auf der nächsten Ebene erhält er dann ein größeres Vorschaubild Hyperlinks für die Auswahl der lesbaren Ansicht.

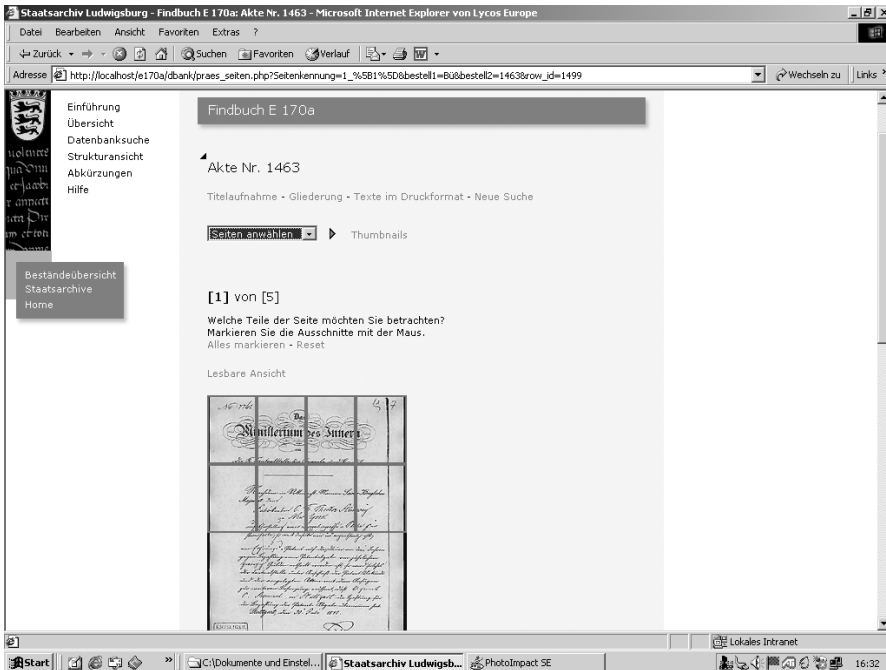


Abb. 13: MIDOSAonline Findbuch. Vorschaubild eines Dokuments

Aufgrund der geringen Bandbreite des Internets wurde zur Reduktion der Übertragungszeit für die Bilder der lesbaren Ansicht eine besondere Technik entwickelt, die neben einer optimalen Bildqualität rasche Übertragungszeiten ermöglicht. Die einzelnen Bilddokumente wurden "geka-chelt" bzw. segmentiert, so dass immer nur die über das Vorschaubild ausgewählten Bildteile geladen werden.

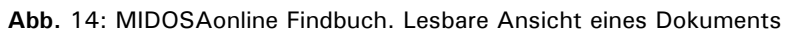


Abb. 14: MIDOSAonline Findbuch. Lesbare Ansicht eines Dokuments

Edgar Büttner: Das Archivverwaltungssystem BASYS als Voraussetzung für die Internetpräsenz des Bundesarchivs

I. Einleitung

Das Schreiben eines verwirrten Menschen, der im Jahr 1888 behauptete, der Whitechapel-Mörder zu sein, und Unterlagen über Versuche mit Gasmasken für Hunde im Zweiten Weltkrieg – das, liebe Kolleginnen und Kollegen, sind die beiden Highlights im April, zu finden in jener Liste, mit der das Public Record Office in seiner vorzüglichen Internet-Präsentation auf sich und auf benutzbar gewordene Dokumente Monat für Monat aufmerksam macht, worunter sich – archivüblich - selbstverständlich auch weniger Sensationelles findet. Akten zur ENIGMA-Chiffriermaschine sowie zur Inhaftierung von Rudolph Hess im Spandauer Gefängnis, ebenfalls im monatlichen Angebot enthalten, konnten sich nicht als Highlights platzieren.

Zweifellos, die Dokumente selbst sind historisch seriös, und greifen mit „Jack the Ripper“ und einem Thema aus dem Zweiten Weltkrieg, der im britischen Fernsehen wöchentlich neu gewonnen wird, Gegenstände auf, die sich hoher Beliebtheit erfreuen - doch bleibt die Frage, welche Qualitätskriterien für Highlights zugrundegelegt werden: Einerseits - macht allein die historisch geronnene Sensation den Reiz aus, vielleicht auch ein wenig kriminelle oder kriegerrische Grausamkeit? Andererseits - sind historisch überlieferte Grausamkeiten, etwa in Form von Gasmasken für Tiere, angesichts tagesaktueller Bilder aus englischen und anderen Schlachthöfen auf dem medialen Markt wirklich „wettbewerbsfähig“? Verzerrt eine solche Auswahl von Schlagzeilen nicht den von Archiven erfüllbaren Erwartungshorizont, und bleibt eine solche Verzerrung gänzlich ohne Rückwirkung auf die Wahrnehmung der Quellen bei der Überlieferungsbildung?

Die Aktenfreigaben im Bundesarchiv verlaufen jedenfalls gänzlich unspektakulär, lautlos durch Fristablauf. Jedes Jahr werden erhebliche Aktenmengen gemäß Bundesarchivgesetz benutzbar. Monatliche Nachrichten über „recent releases“ mit Eventcharakter für die Öffentlichkeit fehlen notgedrungen in der Internet-Präsentation. Dies heißt allerdings nicht, dass es nicht etwas gäbe, worauf aufmerksam zu machen, was im guten Sinne zu vermarkten sich lohnt. Archive leben letztlich nicht von der Schlagzeile, sondern von der Zuverlässigkeit und Seriosität, mit der sie Informationen zur Verfügung stellen.

II. Ziele der Internet-Präsentation des Bundesarchivs

Ein Teil von Ihnen wird die derzeitige Internet-Präsentation des Bundesarchivs kennen, in deren Mittelpunkt ein Beständebaum steht. Diese erste seit 1997 bestehende Internet-Präsenz wird in Kürze enden, da das Bundesarchiv zusätzlich zur inhaltlichen auch die gesamte technische Betreuung in die eigenen Hände übernehmen wird. Ein Internet-Server ist angeschafft; er kann jedoch erst später als zunächst beabsichtigt ans Netz gehen, da mit dem Schritt der Übernahme in die eigene Verantwortung auch eine Umstellung des Betriebssystems von WINDOWS NT auf LINUX erfolgen wird. Der notwendige Auftrag ist erteilt, mit dem Abschluss der Arbeiten ist Anfang Juni 2001 zu rechnen. Parallel zur technischen Umstellung hat eine unter Leitung des Präsidenten des Bundesarchivs eingesetzte Projektgruppe „Internet/Intranet-Koordinierung“ die Aufgabe, die Präsentation aus fachlicher Sicht weiterzuentwickeln. Benutzerorientierung sowie Zugang zu den vorhandenen und laufend entstehenden Erschließungsleistungen sind dabei die leitenden Zielvorstellungen.

Damit dieses Angebot konsequent verbessert wird, verfolgt das Bundesarchiv - in der Reihenfolge ihrer Priorität - diese Ziele:

1. Eine Online-Beständeübersicht soll ins Internet gestellt werden,
2. Online-Findbücher zu ausgewählten Beständen sowie eine Online-Datenbank-Replikation und
3. ausgewählte digitalisierte Dokumente sollen folgen.

Wichtige inhaltliche Schritte zur Umsetzung des ersten Ziels sind bereits geleistet: die Beschreibungen für die Bestände der Abteilung Bundesrepublik liegen vollständig vor, die bereits in Buchform veröffentlichten Beschreibungen für die Bestände der Abteilung DDR in einer aktualisierten Version; Beschreibungen der maschinenlesbaren Datenbestände sind ebenfalls vorhanden. Der nächste Schritt wird die Präsentation der Beständeübersicht mithilfe von MIDOSA-Online sein. Die hierzu erforderlichen Daten sollen aus den Stammdaten über Provenienzen und Bestände der Datenbank des Bundesarchivs automatisch ausgelesen werden. Da die benutzerorientierte Präsentation von Erschließungsleistungen im Internet eine vorrangige Aufgabe aller Archive ist, wird das Bundesarchiv neben den notwendigen eigenen Anstrengungen daher gerade auf diesem Gebiet auch diejenigen Möglichkeiten und Werkzeuge nutzen, die Projekte und Kooperationen wie die „Partnerschaft internetfähige Archivlösungen“ PARSIFAL bieten.

III. Das Archivverwaltungssystem BASYS

1. Ausgangslage des Bundesarchivs

Damit die Präsentation von Erschließungsinformationen über Anfangserfolge hinaus aktuell und erfolgreich bleiben kann, muss eine effiziente Gewinnung oder leistungsfähige Produktion vorausgesetzt werden. Der Umfang der im Bundesarchiv verwahrten Überlieferungen ist beachtlich, im Jahr 2000 betrug er 280.862 lfm. Davon entfielen auf die „historischen Abteilungen“ Reich und DDR 36.388 bzw. 42.291 lfm Archivgut; signifikante Zuwächse sind in diesen Bereichen nicht mehr zu erwarten.

Die „Wachstumsbranche“ des Bundesarchivs hingegen ist zweifellos die Abteilung Bundesrepublik mit ihren Ministerialbeständen; mit einem Sprachbild aus dem weltwirtschaftlichen Kontext könnte man sie als „tiger division“ bezeichnen. Im Jahr 2000 befanden sich in ihrer Zuständigkeit 98.462 lfm Unterlagen und Archivgut.¹ Jährlich gelangen in die Zwischenarchive in St. Augustin Hangelar und Dahlwitz-Hoppegarten, die für die Übernahme der zivilen Ministerialbestände zuständig sind, etwa 90.000 bis 100.000 Aktenbände, das entspricht im Durchschnitt der letzten Jahre ca. 6.000 lfm².

Auch der Umfang einzelner Ministerialbestände ist beachtlich, denn ein Bestand umfasst alle bei einer Provenienz angefallenen Akten, nicht allein Abgabeschichten: für die „großen“ unter ihnen ist mit Zuwächsen zwischen einem halben und einem Kilometer im Jahr zu rechnen³. Derzeit (Stand: 4. Mai 2001) beträgt der Umfang des Bestands Bundesmi-

¹ In die Zahlen nicht einbezogen sind die Bestände der Abteilung Militärarchiv aus den Epochen Reich, DDR und Bundesrepublik.

² Die gerundeten Gesamtzahlen für die letzten drei Jahre lauten:

1998	6.200 lfm
1999	7.500 lfm
2000	5.700 lfm.

³ Bezogen auf die Bestände Bundesministerium des Innern (B 106), Bundesministerium der Finanzen (B 102) und Bundesministerium für Wirtschaft (B 102) lauten die Zahlen z.B.:

	B 106	B 126	B 102
1998	589 lfm	482 lfm	1.180 lfm
1999	1.541 lfm	778 lfm	581 lfm
2000	1.064 lfm	786 lfm	298 lfm.

nisterium der Finanzen B 126 beispielsweise 149.545 Akten (oder 7.777 lfm), der des Bestands B 106 Bundesministerium des Innern 148.156 Akten (6.518 lfm) und der des vergleichsweise kleinen Bestands Bundesministerium der Justiz B 141 149.387 Akten (oder 2.220 lfm). Nur ein Teil dieser Unterlagen ist archivwürdig; aus unterschiedlichen Gründen, die auch in der Schriftgutverwaltung liegen, schwankt der prozentuale Anteil: im Falle des Bundesministeriums der Justiz sind 55%⁴, im Falle des Bundesministeriums des Innern 15-20% archivwürdig. Wenngleich daher Erschließungsinformationen nur zu einer Teilmenge der „zwischenarchivischen Grundgesamtheit“ anfallen, bleiben auch die reduzierten Zahlen erheblich. Die aktuelle Anzahl der archivwürdigen – großteils aus der Zeit bis ca. 1970 stammenden - Akten lautet z.B. für folgende Bestände (Stand: 1. Februar 2001):

Bundesministerium für Wirtschaft	B 102	26.750,
Bundesministerium der Justiz	B 141	26.596
Bundesministerium des Innern	B 106	21.563
Bundesministerium für Arbeit und Sozialordnung	B 149	17.131
Bundesministerium für Bildung und Wissenschaft	B 138	16.455
Bundesministerium der Finanzen	B 126	12.541

Geschichte und Erfolg des Bundesarchivs sind mit dem Zwischenarchiv⁵, d.h. mit seiner intellektuellen Konzeption und den praktischen Verfah-

⁴ Das BMJ verfügt über eine effiziente Zentralregistratur, die bereits bei der Entstehung der Akten die aus dem bürokratischen Informations- und Mitwirkungsprozess resultierende Redundanz von Informationen erheblich reduziert.

⁵ Heinz Hoffmann: Behördliche Schriftgutverwaltung. Ein Handbuch für das Ordnen, Registrieren, Aussondern und Archivieren von Akten der Behörden, Boppard am Rhein 1993 (Schriften des Bundesarchivs Band 43), S. 363-368.

rensabläufen⁶, eng verbunden. In diesem Kontext wurde ein kombiniertes Verfahren für die *Bewertung und Erschließung* der Ministerialakten⁷ entwickelt. Vor dem Hintergrund der genannten Zahlen ist die technische Umsetzung dieses Verfahrens nicht allein auf die Erschließung eines einzelnen Aktenbandes über einen einzelnen Datensatz ausgerichtet; vielmehr ist auch die Bearbeitung größerer Mengen von Akten (Selektionen) in einem Arbeitsschritt notwendig. Die Bearbeitung von Selektionen ist daher bei verschiedenen Verfahrensschritten im Bundesarchiv vorgesehen: bei der Bewertung von zu vernichtenden Akten wie auch bei der Zuordnung archivwürdiger Akten zu einem Klassifikationsschema, das sich möglichst an vorarchivische Ordnungsstrukturen anlehnen wird.

Welche Folgerungen für die Präsentation von Erschließungsinformationen sind daraus zu ziehen?

Die traditionelle archivarisches Zielkategorie „Findbuch“ – in ihrer modernen Gestalt des Online-Findbuchs – behält sicherlich ihre Berechtigung⁸.

⁶ Siehe Gemeinsame Geschäftsordnung der Bundesministerien: Allgemeiner Teil (GGO I), Besonderer Teil (GGO II) [...] Hrsg. Vom Bundesministerium des Innern. Stuttgart 1974 ff., Loseblattausgabe).

⁷ Siehe zuletzt Wolfram Werner: Zur Überlieferungsbildung im Bereich des staatlichen Schriftgutes der Bundesrepublik Deutschland. Versuch einer Zwischenbilanz, in: Archiv und Geschichte: Festschrift für Friedrich P. Kahlenberg, Düsseldorf 2000 (Schriften des Bundesarchivs Band 57), S. 348-364.

⁸ Mechthild Black-Veldtrup: Erschließung im Umbruch, in: Der Archivar 1998, Sp. 607-618; Dies.: Findbücher im Internet, in: Angelika Menne-Haritz (Hg.): Archivische Erschließung, Methodische Aspekte einer Fachkompetenz, Marburg 1999, S. 123-138; Dies. und Detlef Heiden: Das Marburger Online-Findbuch. Konsequenzen für die Erschließung und Präsentation von Archivgut, in: Der Archivar 1999, S. 217-224; Paul Flamme, Udo Herkert und Volker Viergutz: Hinweise zur Darstellung staatlicher Archive und Archivverwaltungen im WorldWideWeb des Internet, in: Der Archivar 1998, Sp. 217-228; Angelika Menne-Haritz: Einige Überlegungen zur Interaktivität von Archivfindmitteln, in: Der Archivar 49, 1996, Sp. 603-610; Dies.: Das Online-Findbuch - archivische Erschließung mit

Dieses Ziel wird auch im Bundesarchiv nach wie vor bei abgeschlossenem, d.h. in der Regel historischen Beständen anzustreben sein. Dies gilt insbesondere für Bestände von hohem Wert und hoher Benutzungsrelevanz wie etwa Reichskanzlei oder Reichsjustizministerium. Das gleiche gilt für Teile von Ministerialbeständen der Bundeszeit, wie etwa für die Akten der Kulturredaktion des Bundesministeriums des Innern, nicht aber in gleicher Weise für die Akten der Abteilungen Bundesgrenzschutz oder Dienstrecht.

Es ist offensichtlich, dass vor dem Hintergrund eher schwindender Ressourcen bei weitem nicht alle in der Datenbank des Bundesarchivs enthaltenen Erschließungsinformationen in die Form hergebrachter Findbücher mit ihrer ausgeklügelten und am Druckbild orientierten Sortierung gebracht werden können. Und es ist eine Erfahrung, dass perfektionistische Findbücher mit wenigen Seiten leichter als umfangreiche mehrbändige Findbücher entstehen. Das kann aber nicht heißen, dass die genannten Informationen für Recherchen vernachlässigbar wären. Wenn der Zugang zu allgemein benutzbaren Überlieferungen zentrales archivpolitisches Ziel ist, besteht vielmehr die Notwendigkeit, auch solche Erschließungsinformationen, die sonst nur in der Datenbank vorhanden wären, auch im Internet zu präsentieren.

Für Überlieferungen, bei denen laufend ein Zuwachs an Archivgut und damit an Erschließungsinformationen anfällt, bedarf es daher der Ergänzung durch eine möglichst tagesaktuelle Datenbank-Recherche, die dem dynamischen Charakter der archivischen „Wachstumsbranche Ministerialakten“ angemessen ist und die Chancen der neuen Medien voll nutzt.

2. Historie: AKABDA und das Zwischenarchiv

Seit Ende der 70er Jahre bis 1992 war im Bundesarchiv das System AKABDA in Betrieb. Die Abkürzung stand für „Aktenabgabedatei“ und erklärt sich damit weitgehend selbst. Ihr primärer Zweck lag in der zwischenarchivischen Aufgabenstellung; das sind: Verwaltung der zur Zwischenarchivierung übernommenen Unterlagen, Überwachung der Aufbewahrungsfristen, Eintrag der Bewertungs- und Erschließungsergebnisse, Umsetzung der Bewertung nach Fristablauf, Durchführung der Kassationen. Der Einsatzbereich von AKABDA war, gesehen für das gesamte Bundesarchiv, ein - wenn auch bedeutsamer - Sonderbereich: nämlich Schriftgut aus der Zeit der Bundesrepublik und hier wiederum nur Schriftgut der Ministerien. Bei der Weiterverarbeitung der Erschließungsinformationen zeigte AKABDA deutliche Schwächen; zwar ließ sich eine Klassifikation herstellen, doch war es nur mit einiger Bastelei möglich, Findbücher auszugeben.

3. Das IT-System BASYS

Angesichts seiner begrenzten Einsatzmöglichkeiten und seiner Schwächen musste AKABDA einem umfassenden Archivverwaltungs- und Erschließungssystem weichen. 1988 wurde mit einem fachlichen Feinkonzept für ein IT-System begonnen. Das Projekt war anspruchsvoll und die Realisierung zog sich über mehrere Jahre hin. Das IT-System ging 1992/1993 in den Wirkbetrieb. Die längste Zeit hat das Bundesarchiv mit einem No-Name-Produkt gearbeitet – und das in einer Zeit, in der sich Markenbewußtsein als ein Mentalitätsmerkmal in der bundesrepublikanischen Gesellschaft etabliert hat, bis hinein in die hintersten Winkel der Kinderzimmer. Mittlerweile nennen wir das System BASYS, d.h. Bundesarchiv-IT-System, und der Anklang daran, daß es unser grundlegendes Arbeitswerkzeug ist, ist so gewollt.

BASYS wurde konzipiert als ein integriertes Archivverwaltungssystem, das über verschiedene Programmbereiche verfügt, die den Bearbeitungs-
zustand eines archivischen Objekts abbilden, von der Übernahme ins
Zwischenarchiv oder Endarchiv bis hin zur Benutzung.

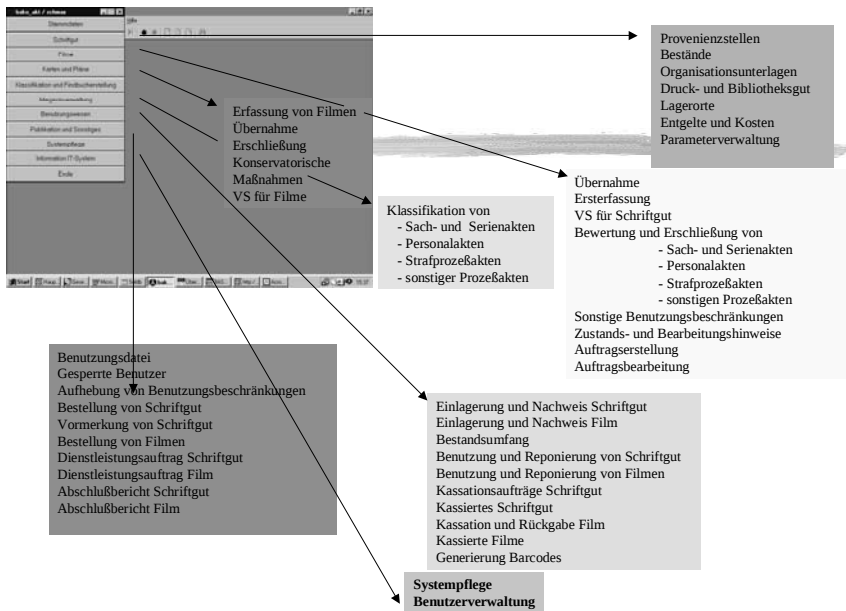


Abb.1: Startmenü

Es besteht aus zwei Teilsystemen, dem Bereich „Schriftgut“ und dem Bereich „Film“, den ich, um im Bild zu bleiben, im Kontext dieser Darstellung gleich wieder ausblende.

Der Bereich „Schriftgut“ verfügt über folgende Komponenten:

- Informationsstrukturen zur Verwaltung der Bestände – die Tabellen Provenienzen, Bestände und deren Systematisierung sowie Lagerungsorte

- Informationsstrukturen zur Verwaltung von Zugängen – die Tabellen Abgabe und Zugang
- Informationsstrukturen zur Erschließung – die Tabellen Bewertung und Erschließung, Enthält-Vermerke, Zusatzinformationen zu Personalakten, Klassifikation und Findbuch
- Informationsstrukturen zum Bereich der Magazinverwaltung und der Benutzung – Tabellen Magazinverwaltung, Zustands- und Bearbeitungshinweise, Kassationsaufträge, kassiertes Schriftgut, Benutzer, Bestellungen.

4. Derzeitige Überarbeitung des Bereichs Schriftgut

Wer die Entstehung eines IT-Systems betreut, weiß, dass ihm eines sicher ist: Kritik. So wurden an den Anwendungsprogrammen des Bereichs „Schriftgut“ verschiedene tatsächliche oder auch empfundene Mängel beklagt und Verbesserungen gefordert. Dies führte schließlich dazu, dass im Jahr 1999 mit einem neuen fachlichen und DV-technischen Feinkonzept begonnen wurde. Die technische Umsetzung erfolgte während des letzten Jahres durch die Firma PRO DV (Dortmund); Teile der o.g. Informationsstrukturen wurden überarbeitet; neben einer Verbesserung des Anwendungskomforts sollte nicht zuletzt die Nutzbarkeit der Daten erhöht werden. Wenngleich die Testphase noch anhält und deutlich wurde, dass einige der gesetzten fachlichen und technischen Ziele nur mit Verzögerung erreichbar sind⁹, will ich wesentliche Ziele kurz skizzieren:

1. Es ist ein langfristiges Ziel, die zwischenarchivischen Daten von den endarchivischen Daten zu trennen. Ein wichtiger Schritt hierfür ist mit der Einführung einer sog. Info-Datei für den Bereich der „Verwaltung von

⁹ Mit einem breiten Wirkbetrieb ist zur Jahreswende zu rechnen.

Zugängen“ getan. Es handelt sich um eine Tabelle, in die elektronische Abgabeverzeichnisse eingelesen werden können. Die bereits seit mehreren Jahren praktizierte Übernahme maschinenlesbarer Abgabelisten wird damit zu einem standardisierten Verfahren, das aus fachlicher wie aus wirtschaftlicher Sicht gewinnbringend sein wird¹⁰:

Vorarchivische Registraturdaten können damit übernommen und im Zuge der Bewertung zu Erschließungsdaten weiterverarbeitet, Bewertungsentscheidungen können transparent gehalten werden. Mit dieser Komponente wurde ein wesentlicher Kern der Konzeption von AKABDA wieder aufgenommen.

Die „Übergabe“ bezeichnet in diesem Zusammenhang eine Funktion, mit der Erschließungsdatensätze von der Info-Datei in die sog. Prozessdatei kopiert werden. Handelt es sich um massenhafte kassable Unterlagen, kann eine Massenkassation bereits außerhalb des Erschließungsprogramms ausgelöst werden.

Sofern keine elektronischen Abgabelisten vorliegen, können die Informationen zu Aufbewahrungseinheiten manuell erfasst werden. Auch ein externes Erfassungsprogramm ist vorhanden, das Behörden oder auch privatrechtlich organisierten Institutionen zur Verfügung gestellt werden kann.

¹⁰ Siehe hierzu Axel Schmidt: Übernahme und Weiterverarbeitung elektronischer Abgabeverzeichnisse in den Zwischenarchiven. Manuskript eines Vortrags, gehalten anlässlich eines Seminars „Bestandsergänzung - Bewertung und Übernahme in den Archiven Deutschlands und Russlands“ in Seelow. Koblenz, Juni 2001.

BASYS-S - Externe Erfassung

Datei Bearbeiten Funktionen Module Fenster ?

Aufbewahrungseinheiten 1

	Kennung	Lfd.Nr.	Vgg.Nr.	Az.	Inhalt	Lfz. von	Lfz. bis	Bandnr.	Dauer	Ende	Org.Einheit	VS
100	1	1	400	Strafverf	1955	1956	1	30			Z 1	
100	2	1	400	Strafverf	1956	1957	2	30			Z 1	
100	3	1	400	Strafverf	1957	1958	3	30			Z 1	
100	4	1	401	Rechtsgu	1957	1957	1	30			Z 1	
100	4	2	401	Rechtsgu	1960	1960		30			Z 1	
100	5	1	401	Rechtsgu	1958	1959	2	30			Z 1	
100	6	1	400	Zustand	1970	1972	1	30			Z 2	

Aufbewahrungseinheit 1.1

Abgabekennung:

Lfd. Nr.: Vorgangsnummer:

Aktenf., Org.-Einheit:

Aktenzeichen:

Inhalt:

Bandnummer:

Laufzeit: -

Aufbewahrungsfrist: - Dauer: - Endjahr:

VS: Offenlegungsjahr:

Drücken Sie F1, um eine Hilfe zu erhalten.

Start BASYS-S Microsoft PowerPoint - [ma...] BASYS-S - Externe Er... 15:27

Abb. 2: Externe Erfassung

2. Allgemein sollte die Benutzeroberfläche an die derzeitigen technischen Standards angepasst werden; dazu gehört u.a. die volle Windows-Kompatibilität. Mit einer spürbaren Verbesserung des Bearbeitungskomforts sollte zugleich die Akzeptanz der IT-gestützten Erschließung auch in den historischen Abteilungen gesteigert werden. Zu diesem Bereich gehört auch der Texteditor mit seinem bisher nur geringen Textverarbeitungskomfort in den sog. Stammdatenprogrammen, mithilfe derer Informationen über die Provenienzstellen und Bestände gesammelt und gepflegt werden sollen.

3. Gab es bisher nur den Einstieg über den einzelnen Datensatz, sollen nun Listenfenster den Bearbeitungskomfort erhöhen. Die Reihenfolge und Breite der Spalten der Listenfenster kann dabei individuell abgespeichert werden; selbstverständlich kann auch individuell sortiert werden.

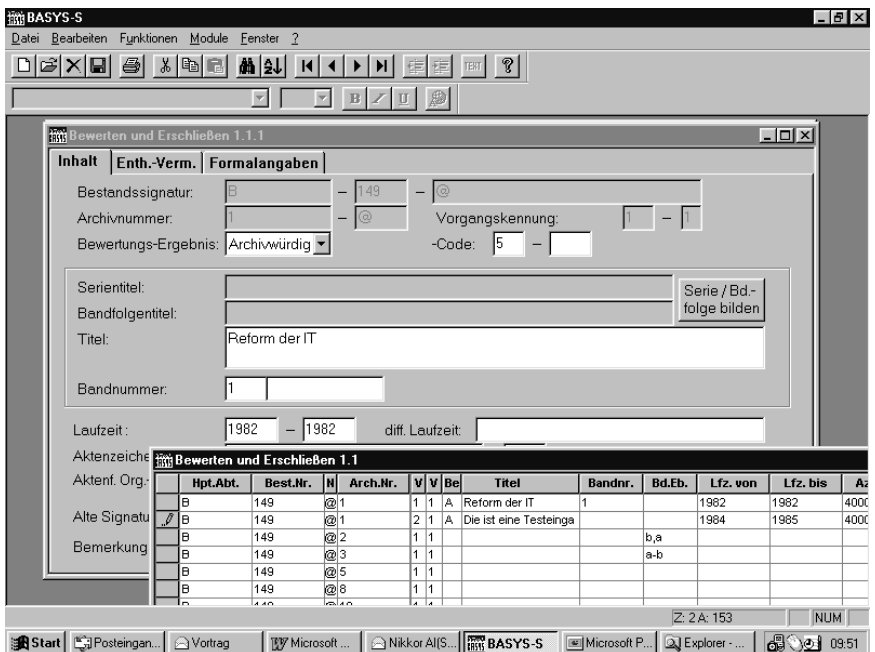


Abb. 3: Detail- und Listenansicht bei der Erschließung

4. Wird in einem umfangreichen Bestand ein Nummernbereich, z.B. die Nummern 1000-1100, bewertet und erschlossen, ist es wichtig, auch einen sachlich gleichen oder ähnlichen Bereich ansehen zu können (z.B. die Nummern 36.000-36050), ohne die aktuelle Auswahl zu verlassen; daher sollen Programme mehrfach aktivierbar sein.

5. Neben Anpassungen der Benutzeroberfläche ist es im Bereich „Erschließung“ das Ziel, die Struktur der archivischen Daten bei Serienbildung und Klassifikation zu verbessern:

- Die Bildung von Bandfolgen und Serien soll durch einen Serienmanager erleichtert werden. Zu diesem Zweck wurde eine neue Tabelle eingerichtet. Die Redundanz der Textinformation „Serientitel“ fällt damit weg.

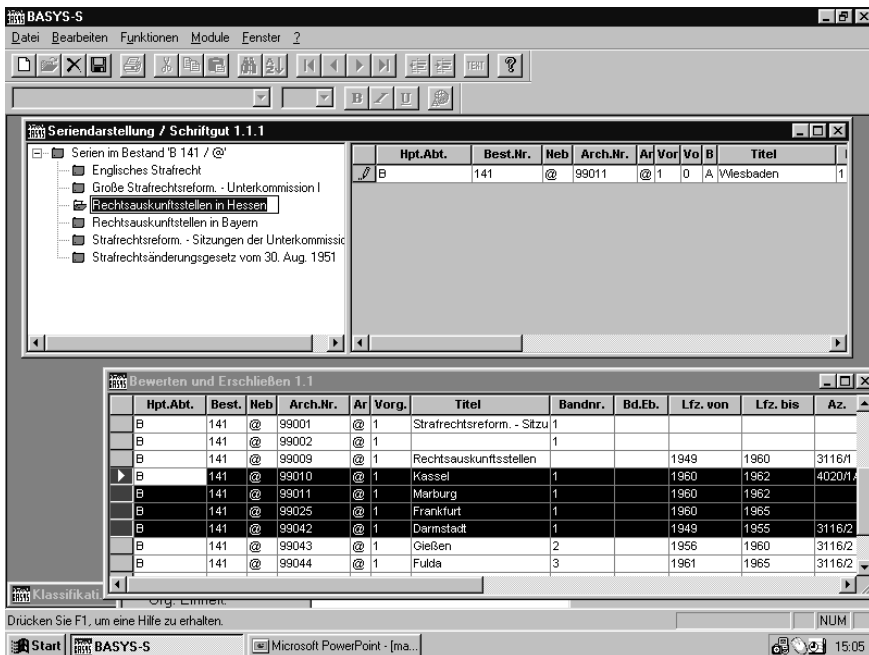


Abb. 4: Serienmanager

- Die bisher getrennten Programme für Klassifikation sowie Bewertung und Erschließung sollen integriert benutzbar sein,
- die Pflege des Klassifikationsschemas und die Zuordnung von Datensätzen zu diesem Schema sollen durch einen Klassifikati-

onsmanager entscheidend verbessert werden (u.a. fensterübergreifende Drag-and-drop-Funktion), der Klassifikationsmanager kann auch beständeübergreifend z.B. für sachthematische Inventare eingesetzt werden,

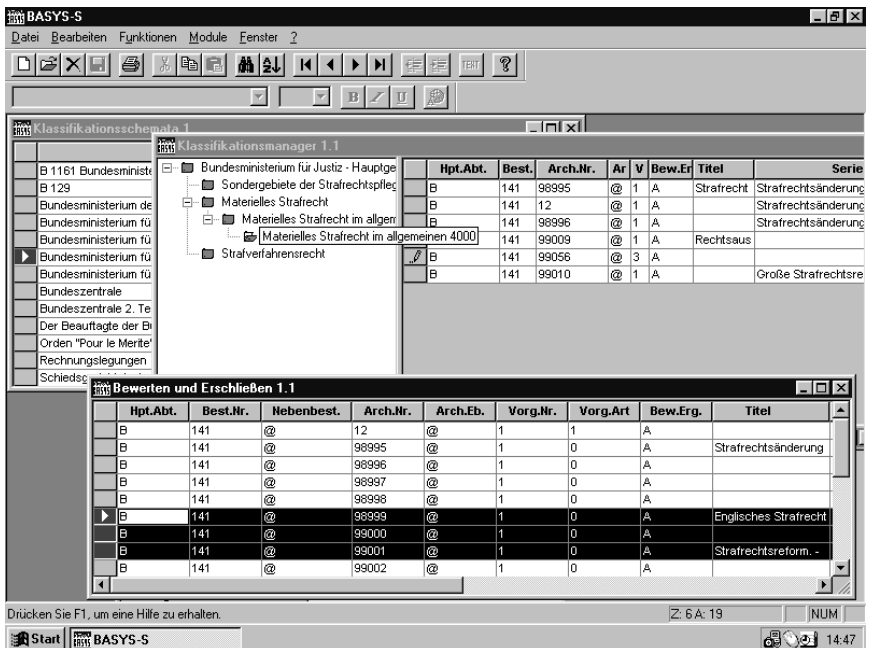


Abb. 5: Klassifikationsmanager

Ein sog. „Rohfindbuch“, das eine strukturierte Sicht auf mehrere Datensätze gibt, soll die Redaktionsarbeiten am Findbuch rationeller gestalten.

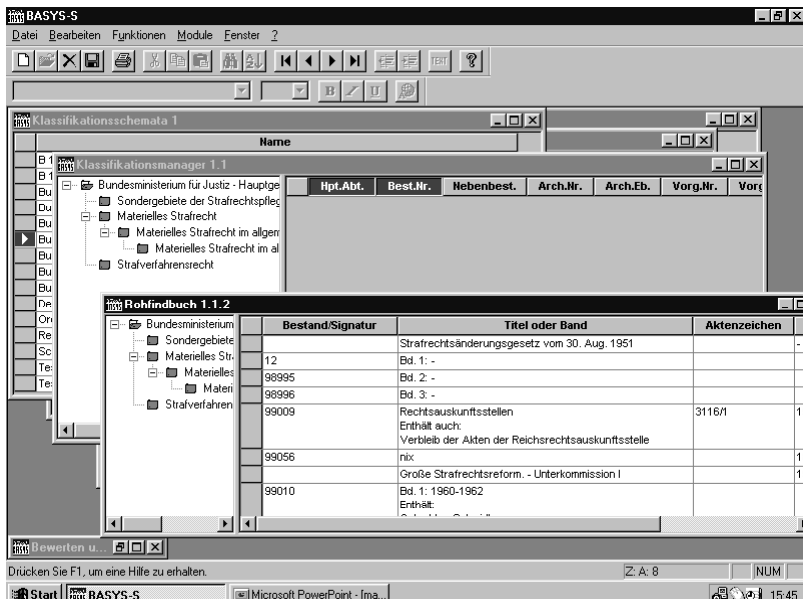


Abb. 6: Rohfindbuch

6. Für die Präsentation von Erschließungsdaten sieht das Feinkonzept vor dem Hintergrund der oben geschilderten Ausgangslage einen zweifachen Weg vor: Beabsichtigt ist einerseits die Findbuchausgabe über verschiedene Schnittstellen, eine RTF-Schnittstelle für den konventionellen Ausdruck und eine XML-Schnittstelle zur Vorbereitung von Online-Findbüchern. Der Findbuchgenerator von MIDOSA-Online gibt dabei den Standard vor. Vorgesehen ist andererseits aber auch die Präsentation von Erschließungsdaten im Internet über eine sog. Datenbank-Replikation.

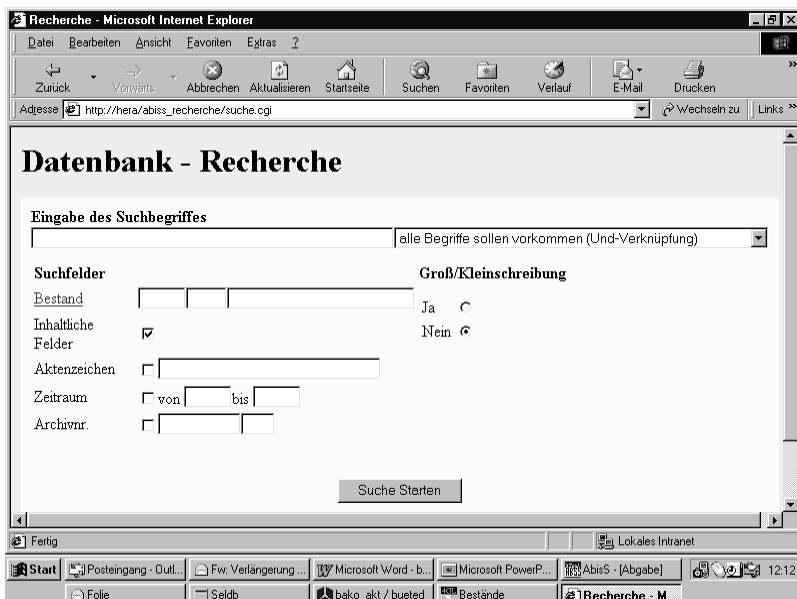


Abb. 7: Datenbank-Recherchemaske

Es handelt sich dabei um ein im Vergleich zu den Navigationsmöglichkeiten von Online-Findbüchern „simples“ Recherche-System, das in der ersten Umsetzung noch keinen Zugriff auf vorhandene Klassifikationsstrukturen und damit auf den Entstehungszusammenhang ermöglicht. Es ist bewusst, dass an dieser Stelle im nächsten Schritt nachgebessert werden muss. Die Ergebnisseite soll, wie dies üblich ist, die Treffer in Form einer Liste anzeigen; von dort wird es dann möglich sein, die Detailansicht eines einzelnen Datensatzes aufzurufen.

Gute Vorsätze allein verändern nicht die Wirklichkeit. Ich will daher auch eine kritische Zwischenbilanz ziehen. Während der bisherigen Testphase traten, wie nicht anders zu erwarten, Fehler und Probleme auf. Verbes-

serungsbedarf besteht jedoch nicht allein bei der Software, sondern auch bei der technischen Infrastruktur; denn die Software reagiert sensibel auf Probleme, die serverseitig bestehen. Lösungen für beide Problembereiche sind in Sicht.

5. Fortentwicklung des IT-Systems

BASYS steht für ein modulares System. Und es besteht Bedarf nach Fortentwicklung. Drei Eckpunkte will ich kurz bezeichnen:

- a) Die bisherige *Film-Datenbank* soll im Laufe dieses Jahres durch ein sog. Standard-Softwareprodukt abgelöst werden; es handelt sich dabei um das System MAVIS der australischen Firma Wizard. Vorgesehen sind Anpassungen und Erweiterungen, die für die Anwendungen im Bundesarchiv erforderlich sind. Der Vertragsabschluß steht unmittelbar bevor.¹¹
- b) Die bisher in Betrieb befindliche Software für die *personenbezogenen Unterlagen* des ehem. Berlin Document Center kann mittelfristig nicht mehr gepflegt werden. Erste Planungen für ein neues Teilsystem haben begonnen, das allerdings nicht allein die alte Software ersetzen, sondern auch die Erschließung anderer personenbezogener Überlieferungskomplexe im Bundesarchiv (v.a. Akten der Zentralen Stelle Ludwigsburg) integrieren soll.
- c) Eine zukünftige Überarbeitung der Stammdatenprogramme wird auch neben der inhaltlichen die technische Beschreibung maschinenlesbarer Bestände angemessen berücksichtigen müssen.

¹¹ Der Wirkbetrieb ist für Ende 2001 vorgesehen.

IV. Schluß

Trotz Gasmasken für Hunde, BSE und MKS und der damit verbundenen Grausamkeiten werfe ich gerne noch einen Blick auf den britischen Pragmatismus und seine Ergebnisse, hier in Form der Website des Public Record Office. Sie ist zweifellos sehr gelungen und leistungsfähig: Inzwischen stehen den Nutzern über einen Online-Katalog Erschließungsinformationen zu mehr als 8,9 Millionen Verzeichnungseinheiten zur Verfügung. Dem PRO ist damit eine pragmatische Anwendung auf der Grundlage der Encoded Archival Description (EAD) gelungen, die nicht von einzelnen Findbüchern ausgeht, sondern alle in der Inventory List Data Base enthaltenen Erschließungsinformationen ins Internet stellt. Auch der Kontext der Einzelinformation wird in der neuen Version der Ergebnisanzeige angemessen berücksichtigt.

Die Präsentation des PRO setzt Maßstäbe für die Präsentation massenhafter Daten, sie setzt jedoch auch große Investitionen und zielsicheren Ressourcen-Einsatz voraus. Im verantwortlichen Information&Records Department wurden zur Vorbereitung des Projekts AD 2001 archivfachliche und IT-technische Kompetenzen gebündelt, unter der Leitung eines Archivars¹²; die ca. 50 Mitarbeiter, mehr als 10% der Mitarbeiter des PRO, wurden von externen „consultants“ unterstützt, die Energien konzentriert, um wirklich alle Findmittel im Internet zugänglich zu machen, mit externer oder auch interner Datenerfassung.

Am Ende will ich jedoch nicht unrealistisch werden, weil das vermeintlich Bessere stets außerhalb der eigenen Wirklichkeit zu wohnen scheint. Denn auch Archive haben, ähnlich wie ihre Gegenstände, gewachsene Strukturen, und Vorsicht ist gegenüber ausschnittshaften Vergleichen an-

¹² Edgar Büttner: Ein Aufenthalt im Public Record Office, in: Mitteilungen aus dem Bundesarchiv, Heft 2-3, 1998, S. 34-41.

gebracht; die Grundstrategie des PRO, mit Konsequenz und großer Kraftanstrengung ein Ziel zu erreichen, verdient allerdings Nachahmung. Die Bedingungen für einen konsequenten Präsentationskurs sind auch unter den anderen hiesigen Voraussetzungen gut: Die Erschließung der Ministerialbestände liegt im System BASYS vor und der Einsatz der IT-gestützten Erschließung setzt sich auch in den „historischen“ Abteilungen Reich und DDR durch; die Aufarbeitung der ehemals von der Staatssicherheit verwahrten Unterlagen der Reichszeit bietet in dieser Hinsicht einen wichtigen Schritt nach vorne. Doch auch die „IT-Erschließungslücken“ in Form konventioneller Findbücher und Textdateien müssen geschlossen werden; vor diesem Hintergrund kommt dem DFG-Projekt „Entwicklung von Werkzeugen zur Retrokonversion archivischer Findmitteln“ der nordrhein-westfälischen Archivverwaltung, an dem auch das Bundesarchiv beteiligt ist, eine wichtige Rolle zu, trägt es doch zu den Grundlagen einer konsequenten archivischen Internetpolitik bei, indem es den Wissensschatz, der in konventionell hergestellten Findmitteln enthalten ist und vor dem Hintergrund der technischen Entwicklungen Gefahr laufe in Vergessenheit zu geraten, gerade in moderner technischer Umgebung nutzbar machen will.

Im Rahmen eines von der DFG bewilligten Projekts des Bundesarchivs mit dem Titel „Internet-Präsentation von Online-Findbüchern unter Berücksichtigung des EAD-Systems“¹³ bietet sich schließlich die Chance, gerade von dem im PRO gesammelten know-how zu profitieren. Parallel zur konventionellen Datenbank-Replikation sollte auch eine EAD-konforme Online-Datenbank-Replikation ins Auge gefasst werden. Aus der Perspektive des Projekts könnte sich auch ein kritischer und im Er-

¹³ Siehe Hans-Dieter Kreikamp: Europäische Ansätze bei der Präsentation von Online-Findmitteln, in: Nils Brübach (Hg.): Der Zugang zu Verwaltungsinformationen, S. 121-128, insbes. S. 128

gebnis pragmatischer Blick auf die Art und Weise ergeben, wie Archive mit vorarchivischen Informationshierarchien umgehen; ich denke, dass es bei der archivischen Erschließung in Zukunft weniger um die Perfektionierung vorarchivischer Strukturen als um das Herausarbeiten handhabbarer und damit präsentabler Informationsebenen gehen sollte.

Mechthild Black-Veldtrup: Das DFG-Projekt „Entwicklung von Werkzeugen zur Retrokonversion archivischer Findmittel“

Einordnung

Wenn man sich die Internetangebote deutscher Archive ansieht, dann fällt auf, dass sich mittlerweile Hunderte von Archiven mit einer eigenen Homepage präsentieren. Bei näherer Prüfung der Seiten im Hinblick auf die im Internet veröffentlichten Findmittel stellt man dann fest, dass inzwischen zahlreiche Archive ihre Beständeübersichten ins Netz gestellt haben – die Deutsche Forschungsgemeinschaft hat dies in Nordrhein-Westfalen ja auch gefördert.¹ Eine Übersicht über das Angebot an Beständeübersichten im Internet pflegt die Archivschule auf ihren Seiten mit den Links zu deutschen Archiven.²

Mit der Präsentation seiner Beständeübersicht im Internet macht ein Archiv die erste Erschließungsebene der von ihm verwahrten Unterlagen einer breiten Öffentlichkeit zugänglich und reizt damit den Appetit seiner Benutzer: Die wollen natürlich mehr, nämlich auch die Findbücher einsehen und – dies nur in Parenthese – am liebsten natürlich auch gleich die Quellen online recherchieren. An dieser Stelle soll es jedoch nicht um die Frage der Präsentation von Quellen gehen, sondern um die Zugänglichkeit von Findbüchern. Archive, die ihre Beständeübersichten im Internet präsentieren, geben eine Richtung vor, in die sie weitergehen wollen, sie stellen dem Benutzer unausgesprochen in Aussicht: Die Findbücher als die zweite Erschließungsebene der Quellen sollen und werden folgen.

¹ <http://www.archive.nrw.de>.

² <http://www.uni-marburg.de/archivschule/fv61.html>.

Sieht man sich jedoch das bisherige Angebot an Online-Findbüchern an (auch hier ist auf die Link-Seiten der Archivschule Marburg zu verweisen), dann wird schnell deutlich, dass es hier einen Bruch gibt: Im Vergleich zu der Zahl der im Internet präsentierten Beständeübersichten sind bisher nur wenige Findbücher online verfügbar. Das liegt mit Sicherheit nicht an einer restriktiven Politik, es liegt auch nicht daran, dass es nur Findbücher gäbe, die aufgrund von Sperrfristen für eine Internet-Präsentation ungeeignet wären – die meisten Bestände sind nämlich frei benutzbar –, und es liegt auch nicht daran, dass es keine Tools gäbe, die Online-Findbücher aus Datenbanken generieren können. Die Deutsche Forschungsgemeinschaft hat mit MIDOSA-Online ein Softwarewerkzeug zur Generierung von Online-Findbüchern aus der Verzeichnungsdatenbank MIDOSA gefördert, und die meisten deutschen Online-Findbücher sind bisher, soweit ich sehe, mit MIDOSA-Online ins Netz gestellt worden. Inzwischen bieten auch die Firmen AUGIAS-DATA, FAUST und SCOPE Internetschnittstellen aus ihren Verzeichnungsdatenbanken an, und das Verwaltungs-, Erschließungs- und Recherchesystem für Archive (V.E.R.A.), das z. Zt. in Nordrhein-Westfalen entsteht, sowie das hessische System HADIS 2000 sehen die Generierung von Online-Findbüchern vor. Wo also liegt das Problem?

Wenn in einem Archiv daran gedacht wird, Findbücher ins Internet zu stellen, wird die Auswahl zunächst nach inhaltlichen und nach benutzerorientierten Kriterien erfolgen: Besonders gut erschlossene, besonders häufig benutzte, besonders wichtige Bestände sollen als erste präsentiert werden. Eine häufig genannte Forderung größerer Archive ist auch, nicht mit einzelnen, beliebigen Findbüchern zu beginnen, sondern die Präsentation im Internet systematisch und in für Benutzer absehbaren Zeiten anzugehen. Diesen inhaltlichen und benutzerorientierten Kriterien stehen in den meisten Archiven technische Probleme entgegen: Die wichtigen, gut erschlossenen, häufig benutzten Findbücher liegen in aller Regel

nicht im Datenbankformat vor und entziehen sich somit der inzwischen kinderleichten oder „archivarsleichten“ Generierung aus der Datenbank in eine Online-Version. Die weitaus meisten Findbücher in deutschen Archiven liegen statt dessen handschriftlich oder maschinenschriftlich und damit in analoger Form oder – in geringerer Zahl – als Ausdruck einer Textdatei vor. Und das ist das Problem, bei dessen Lösung das Projekt ansetzt, das ich Ihnen hier vorstellen möchte.

Projektziele

Die Deutsche Forschungsgemeinschaft hat zum 1. Mai dieses Jahres ein Projekt zur „Entwicklung von Werkzeugen zur Retrokonversion archivischer Findmittel“ bewilligt, das auf zwei Jahre angelegt und am nordrhein-westfälischen Hauptstaatsarchiv in Düsseldorf als federführender Stelle angesiedelt ist. Es verfolgt mehrere Ziele:

1. Es sollen DV-technische Werkzeuge entwickelt werden, mit denen archivische Findbücher aus einer analogen Form oder aus Textdateien in ein strukturiertes digitales Format konvertiert werden können. D. h. dass es nicht darum geht, im Rahmen des Projekts sämtliche Findbücher z. B. des Hauptstaatsarchivs retrospektiv zu konvertieren, sondern es geht um die Entwicklung von Softwareinstrumenten, die die Retrokonversion automatisch oder weitgehend automatisch durchführen.
2. Aus verschiedenen Retrokonversionsmethoden sollen auf der Grundlage einer breiten Basis möglichst unterschiedlicher Vorlagen im Echtbetrieb Kosten- und Aufwandsschätzungen für die Retrokonversion von Findbüchern durchgeführt werden. Dabei sollen Zeitaufwand, Kosten und technische Voraussetzungen ermittelt werden, um Anhaltspunkte für künftige Langzeitplanungen für eine möglichst flächendeckende Retrokonversion von Findbüchern zu gewinnen.

3. Angesichts der heterogenen archivischen Datenbankwelt sollen die digitalisierten Findbuchdaten in unterschiedliche Verzeichnungsdatenbanken überführt werden können. Dazu bedarf es eines Austauschformats für Findbuchdaten, das es bisher in Deutschland nur in Ansätzen gibt und das im Rahmen des Projekts erarbeitet werden soll.
4. Ziel ist die Bereitstellung von Findmitteldaten in Informationssystemen für einen navigierenden und einen recherchierenden Zugriff, nämlich
 - in Verzeichnungs- und Recherchedatenbanken in den Archiven, mit denen Archivare und Benutzer arbeiten, und
 - in Online-Informationssystemen, die über das Internet zugänglich sind.

An dieser Stelle werden aber auch die Grenzen des Projekts deutlich. Im Rahmen des Projekts selbst sind keine Internetpräsentationen geplant. Es geht vielmehr allein um die Bereitstellung von Findbuchdaten, die jedes Archiv, das die Konversionswerkzeuge nutzt, nach eigenen Zielvorstellungen zugänglich machen kann, d. h. es geht darum, die Voraussetzung für breit angelegte Internetpräsentationen und Internetabfragen von Findbüchern überhaupt erst zu schaffen. Gleichwohl haben sich die beteiligten Archive der DFG gegenüber verpflichtet, die im Rahmen des Projekts konvertierten Findbücher nach dessen Abschluss ins Internet zu stellen.

An dieser Stelle möchte ich auch auf eine weitere Grenze des Projekts hinweisen: Es geht allein um die Digitalisierung von Findbüchern, nicht um deren Neuerschließung, die angesichts der zu konvertierenden Mengen in diesem Rahmen nicht zu leisten wäre.

Kooperation

Die Anlage des Projekts erfordert die Kooperation mehrerer Archive und Archivsparten. Nur so ist die Erstellung eines Austauschformats, das Akzeptanz finden soll, zu bewerkstelligen, nur so lassen sich unterschiedlichste Findbuchtypen bezüglich ihrer Konversionsfähigkeit testen. Die Probleme, die gelöst werden sollen, betreffen alle deutschen Archive in vergleichbarer Weise. Wenn die Lösungen, die im Projekt gefunden werden, tragfähig sein sollen und wenn die Softwarewerkzeuge von möglichst vielen Archiven nachgenutzt werden sollen, muss bei ihrer Entwicklung ein archivspartenübergreifender fachlicher Konsens gefunden werden. Beteiligt sind deshalb das Bundesarchiv als großer Partner, die vier nordrhein-westfälischen Staatsarchive in Düsseldorf, Münster, Detmold und Brühl, das westfälische Archivamt in Münster und das rheinische Archiv- und Museumsamt in Brauweiler, die die Interessen der über 400 Kommunalarchive in NRW bündeln, sowie das Westfälische Wirtschaftsarchiv in Dortmund.

Methoden

Im Folgenden soll die Umsetzung der Projektziele im einzelnen erläutert werden. Zunächst zu den Softwarewerkzeugen:

Die retrospektiv zu konvertierenden Findbücher werden für dieses Projekt allein technisch, nämlich nach der Form ihrer Vorlage unterschieden, inhaltliche Aspekte spielen hier keine Rolle. Archivare pflegen Findbücher in Buchform zu bringen; die Vorlagen können aber sehr unterschiedlich sein. Findbücher liegen in Datenbankformaten, in Textdateien, maschinenschriftlich oder handschriftlich vor. Im Projekt geht es in drei Teilprojekten um die Retrokonversion von handschriftlichen und maschinenschriftlichen Findbüchern und von solchen, die mit Textverarbeitung entstanden sind und bei denen die Textdateien sich erhalten haben. Die Datenbankformate bleiben außer Acht; hier haben sich die beteiligten Ar-

archive verpflichtet, sie mit eigenen Mitteln in die jeweiligen Verzeichnungsdatenbanken zu bringen.

Was die Archive jetzt beginnen, haben die Bibliotheken mit der Retrokonversion ihrer Zettelkataloge seit langem vorgemacht. Aus der Bibliothekswelt stammt denn auch der Ausdruck „Retrokonversion“, der für die retrospektive Digitalisierung steht. Parallelen zu den Findmitteln der Archive bestehen zum einen in der Art der Vorlage: Die Zettelkataloge setzen sich aus handschriftlichen, zum überwiegenden Teil aus maschinenschriftlichen Katalogkarten zusammen. Dies entspricht dem Gros der archivischen Findbücher. Zum zweiten folgen die einzelnen Katalogkarten, ähnlich wie die Titelaufnahmen der Archivalien in den Findbüchern, einer Art Formular, das jedoch keine einheitliche Erscheinungsform aufweist. Die dritte Parallelität besteht in der Anordnung der jeweiligen Einheiten: Die bibliothekarischen Katalogkarten sind in der Regel alphabetisch geordnet, die archivarischen Titelaufnahmen innerhalb einer Klassifikationsstufe sind – kontextgebunden – einer vom Archivar festgelegten inneren Ordnung unterworfen. Darüber hinaus eignet den archivarischen Titelaufnahmen die Zugehörigkeit zu einer Klassifikation, d. h. einer Hierarchie – ein Phänomen, das die linear angeordneten alphabetischen Bibliothekskataloge in dieser Form nicht kennen. Bibliotheken haben gegenüber den Archiven außerdem den Vorteil, mit den Preußischen Instruktionen (PI) bzw. Regeln für die alphabetische Katalogisierung (RAK) zwei Standards zu haben, mit denen die Elemente und deren Anordnung auf Katalogkarten vorgeschrieben sind.

Die Gemeinsamkeiten bei der Frage der Retrokonversion sind insgesamt so groß, dass die Archive für die Retrokonversion ihrer Findbücher auf die Erfahrungen der Bibliotheken zurückgreifen können. Dabei ist jedoch zu berücksichtigen, dass die Bibliotheken bereits in den frühen achtziger

Jahren auf dem damaligen Stand der Technik mit der Retrokonversion begonnen haben.

Die Bibliotheken haben sich im Wesentlichen zweier Verfahren bedient:

1. Abschreiben: Das Gros der Bibliotheken ließ seine maschinenschriftlichen Katalogkarten von kostengünstig arbeitenden Hilfskräften abschreiben oder hat Fremdfirmen mit dieser Aufgabe betraut (z. B. das HBZ in Köln; die Kosten pro Titelaufnahme beliefen sich auf ca. DM 5.-).³ Die Abschrift der handschriftlichen Karten erfolgte dagegen in der Regel durch Diplombibliothekare.⁴
2. Imaging: Vereinzelt wurden die Titelaufnahmen gescannt und als Images online zugänglich gemacht, wobei die Benutzerrecherche durch OCR-Erkennung (Zentralbibliothek in Zürich)⁵ oder durch manuelle Verschlagwortung (Nationalbibliothek in Wien) gewährleistet wurde.

Beide Verfahren sollen auch in dem hier vorgestellten Projekt für die handschriftlichen Findbücher erprobt und hinsichtlich des Aufwands und der Kosten mit automatisierten Lösungen verglichen werden. Denn hier liegt der große Unterschied zu den bibliothekarischen Retrokonversionsmethoden: Die Archive setzen bei der Retrokonversion von Textdateien und von maschinenschriftlichen Findbüchern ganz, bei den handschriftlichen mindestens teilweise auf Softwaretools.

Diese Werkzeuge müssen folgende Qualitäten aufweisen:

1. Findmittel unterschiedlicher Gestalt müssen konvertiert werden können.

³ <http://www.hbz-nrw.de/hbz/retrozk.html>.

⁴ Ebd.

⁵ http://www.eurospider.com/eurospider/solutions/zb_nzz_sol.html.

2. Alle Informationen müssen verlustfrei übertragen und adäquat strukturiert werden können.
3. Die Arbeiten müssen auf einem (fortgeschrittenen) Anwenderniveau ohne Kenntnisse von Programmiersprachen durchgeführt, aber zugleich auch durch Eingriffe auf Programmebene optimiert werden können.
4. Die Arbeit mit einem Retrokonversionssystem muss deutliche Zeiteinsparungen im Vergleich zum Abschreiben von Findmitteln durch eine Schreibkraft erbringen.
5. Das Zielformat der retrokonvertierten Informationen darf nicht proprietär sein.
6. Die im Laufe des Projekts zu entwickelnden Strukturen für das Zielformat müssen flexibel und erweiterbar sein.

Teilprojekt Retrokonversion von Textdateien

Die Ausgangslage sind Findbücher in verschiedenen Versionen von Textverarbeitungsprogrammen, in der Regel Word oder WordPerfect. Bereits bei der Konvertierung in die neueste Version des jeweiligen Textverarbeitungsprogramms können Layoutmerkmale des ausgedruckten Findbuchs verloren gehen. Die Einzelteile eines typischen Findbuchs sind:

- Titelseite
- Einleitung
- Abkürzungsverzeichnis
- Klassifikation
- Titelaufnahmen
- Index

Die einzelnen Teile sind durch Überschriften in der Regel klar voneinander abgegrenzt; die Einleitung stellt sich als ein Fließtext dar, das Abkür-

zungsverzeichnis als eine Liste u. s. w. Die einzelnen Titelaufnahmen von Akten weisen ihrerseits typische Bestandteile wie

- Signatur
- Aktentitel mit Vermerken
- Laufzeit
- und eventuell Aktenzeichen auf,

die einer bestimmten optischen Anordnung unterliegen. Die Layoutmerkmale sind von essentieller Bedeutung, denn sie machen die einzelnen Funktionen der Bestandteile einer Titelaufnahme optisch unterscheidbar. Ein Beispiel: Die Signatur einer Titelaufnahme erscheint möglicherweise in der ersten Zeile links und fett gedruckt, die Laufzeit ebenfalls in der ersten Zeile rechts und ebenfalls fett gedruckt – das Layout ist also nicht dekorativ, sondern ist Ausdruck der Struktur eines Formulars. Das Layout selbst kann auf unterschiedliche Weise entstanden sein: durch die konsequente Nutzung von Formatvorlagen, durch die Verwendung von Tabellen oder, wohl meistens, durch wilde Formatierungen wie Tabulatoren und Leerzeichen. Schriftauszeichnungen wie fett, unterstrichen, kursiv und gesperrt können gewählt worden sein. Zeilenabstände haben Aussagekraft, aber auch Schlüsselbegriffe wie „Enthält“ oder „Darin“ bezogen auf die einzelne Titelaufnahme oder wie beispielsweise „Einleitung“, „Vorwort“, „Register“, „Index“ bezogen auf die Teile eines ganzen Findbuchs. Diese Gliederungselemente helfen normalerweise dem Benutzer im Lesesaal bei der Orientierung im Findbuch. Bei diesem Projekt geben sie Anhaltspunkte für das zu entwickelnde DV-Tool: Das Softwareinstrument muss die Layoutmerkmale erkennen, und es muss sie für das jeweilige Findbuch nutzen, um die Informationen zu identifizieren und zu strukturieren.

Teilprojekt Retrokonversion maschinenschriftlicher Findbücher

Für die maschinenschriftlichen Findbücher gilt ähnliches wie für die Findbücher, die als Textdateien vorliegen. Das Ziel ist dasselbe, auch hier muss das DV-technische Werkzeug die Findbuchinformationen zerlegen, erkennen und zuordnen. Allerdings muss hier ein Arbeitsgang vorgeschaltet werden, die Findbücher müssen nämlich zunächst gescannt und dann mit einer Zeichenerkennungssoftware (OCR/ICR) bearbeitet werden.

An dieser Stelle ein paar Worte zur praktischen Umsetzung dieses Verfahrens: Die beteiligten Archive richten in ihren Häusern Retrokonversionszentren ein. Dazu sind in Vorbereitung des Projekts als Eigenleistungen der Archive leistungsfähige Aufsichtscanner angeschafft worden. In den Retrokonversionszentren werden die maschinenschriftlichen Findbücher digitalisiert. In der Entwicklungsphase und im anschließenden Echtbetrieb während der Gewährleistungsphase sollen mindestens 150 maschinenschriftliche Findbücher konvertiert werden, um die Softwaretools zu testen und sichere Angaben über Kosten und Aufwand zur erzielen. Nach Abschluss des Projekts sollen die Retrokonversionszentren als Serviceeinrichtungen für die anderen Archive im Land weiterbestehen.

Teilprojekt Retrokonversion handschriftlicher Findbücher

Die Retrokonversion handschriftlicher Findbücher stellt den schwierigsten Teil des ganzen Projekts dar. Denn die handschriftlichen Findbücher machen den Löwenanteil unter den Findbüchern überhaupt aus; sie werden außerdem auch am häufigsten benutzt. Und sie stellen uns technisch darüber hinaus noch vor die größten Probleme. Zunächst war auch einmal zu klären, was, technisch betrachtet, unter handschriftlichen Findbüchern verstanden werden muss, nämlich:

- maschinenschriftliche Findbücher mit zahlreichen handschriftlichen Zusätzen
- rein handschriftliche von einer oder mehreren Händen in lateinischer Schrift
- rein handschriftliche von einer oder mehreren Händen in deutscher Schrift (Sütterlin)
- Mischformen der genannten Typen und
- Zettelkataloge.

Für die Retrokonversion dieser Gruppe von Findbüchern bietet sich bisher noch kein Königsweg an, es sollen jedoch im Verlauf des Projekts drei Wege getestet werden.

1. Das Abschreiben der Findbücher in Anlehnung an die bibliothekarischen Erfahrungen.
2. Scannen der Findmittel und Speicherung im Image-Format, dies ebenfalls in Anlehnung an die Verfahren der Bibliotheken. Bei dieser Methode müssen die Erschließungsinformationen, die für einen navigierenden oder einen assoziativen Zugang erforderlich sind, allerdings ergänzt werden, nämlich Klassifikation, Schlüsselbegriffe und Archivsignaturen. Das Verfahren eignet sich allerdings am ehesten für die archiveigene Erschließung; in Archivverbünden sind Images schwer vorstellbar, wenn unterschiedliche Datenbanken miteinander kooperieren sollen.

Die Anwendbarkeit dieser beiden an bibliothekarischen Katalogen erprobten Verfahren auf die archivischen Findmittel soll im Rahmen des Projekts getestet werden, wobei das Interesse sich insbesondere auf den Zeit- und Kostenaufwand für die Nachbearbeitung richtet.

3. Zusätzlich soll jedoch auch bei den handschriftlichen Findbüchern die Entwicklung der OCR/ICR-Erkennung verfolgt werden.

Mindestens eine Firma bietet Werkzeuge zur Mustererkennung handschriftlicher Texte an, die getestet werden sollen.

Ziel dieses Teilprojekts ist es, einen Vergleich der Kosten und des Aufwands bezüglich aller drei Methoden durchzuführen.

Teilprojekt Austauschformat

Wohin nun mit den konvertierten Daten? Da das Projekt nicht e i n Archiv und nicht nur e i n e Verzeichnungsdatenbank im Blick hat und da es die Nachnutzbarkeit der zu entwickelnden Softwareinstrumente gewährleisten will und muss, stellt sich die Frage nach einer Schnittstelle zwischen Erschließungsinformationen und Verzeichnungsdatenbanken allgemein. Ein solches Austauschformat gibt es bisher für deutsche Findbuchdaten noch nicht. Wohl existieren bibliothekarische Vorbilder wie etwa das MAB II-Format und Dublin Core. Im archivischen Bereich gibt es im anglo-amerikanischen Bereich natürlich EAD (Encoded Archival Description), das aber nicht ohne Anpassungen auf deutsche Findbücher angewendet werden kann. Die bisherigen deutschen Ansätze wie die MIDOSA-DTD für EAD-generierte Findbücher oder die Schnittstelle für die Konversion von Findbüchern in die hessische Verzeichnungsdatenbank HADIS 2000 basieren auf Lösungen für jeweils eine einzige Verzeichnungsdatenbank. Im Retrokonversionsprojekt sollen alle bisherigen Ansätze für ein allgemeines Austauschformat genutzt und nach Möglichkeit integriert werden. Das Ziel ist ein breiter Konsens, der das Austauschformat zu einem Standardaustauschformat werden lässt, einer Schnittstelle, die die Hersteller von Verzeichnungsdatenbanken in ihre Produkte integrieren. Dass bei der Erarbeitung auf die Kompatibilität zu EAD zu geachtet wird, ist selbstverständlich: Das Austauschformat soll natürlich auch die Integration von Findbüchern in große internationale Verbünde ermöglichen.

Das Austauschformat soll folgende Qualitäten erfüllen:

1. Es soll die hierarchischen und relationalen Zusammenhänge von Beständen, Archivalien und Einzelinformationen zu Archivalien adäquat, ohne Informationsverlust und dateiübergreifend abbilden können.
2. Es soll in einem möglichst flachen, standardisierten Dateiformat dargestellt werden. Hier bietet sich das ASCII/ANSII-Format an.
3. Informationen müssen mittels Tags definiert und strukturiert werden. In Anlehnung an die amerikanischen Maßnahmen zur Fortentwicklung von EAD sollen die Tags auf der Basis von XML (Extensible Markup Language) definiert werden, das sich inzwischen zu einem Standard entwickelt hat. Grundsätzlich muss der Standard so offen sein, dass sowohl eine Erweiterung als auch der Übergang auf eine andere, zukünftig zu entwickelnde Markup-Language möglich ist.

Fazit

Das Projekt markiert einen ersten Anfang. Das Austauschformat und die Softwaretools schaffen erst die Voraussetzung für die Retrokonversion großer Mengen unterschiedlicher Findbücher. Auch dieses Ziel wird nur mit Drittmitteln erreichbar sein, selbst wenn die Projektergebnisse natürlich nachnutzbar sein werden. Insgesamt eröffnet sich den Archiven hier die Möglichkeit, mittelfristig das Wissen der Archivare über die Bestände und die Akten einer breiten Öffentlichkeit zugänglich zu machen und sich auf diese Weise neue Benutzerkreise zu eröffnen.

Manfred von Boetticher: Kartographische Darstellung historischer Zusammenhänge im archivischen Internetauftritt - Neue Formen einer benutzungsorientierten Suchstrategie

Seit bald 20 Jahren arbeiten die niedersächsischen Staatsarchive bei der Erfassung von Aktentiteln mit dem - inzwischen auf *Windows* basierenden - Programm AIDA.¹ Oberste Klassifikationsebene stellt dabei - wie in Archiven üblich - der Bestand dar. Das auf diese Weise als Ausdruck - oder inzwischen auch virtuell am Bildschirm erzeugte Findbuch ist dem Benutzer zur eigenen Einsichtnahme zugänglich. Welches Findbuch - oder besser: das Findbuch zu welchem Archivbestand - jedoch bei einer bestimmten Fragestellung heranzuziehen ist, ist vom Benutzer in der Regel nicht zu erkennen.

Die Frage, welche Behörde zu welchem Zeitpunkt in welcher Region für welche Verwaltungsaufgaben zuständig war, macht nach wie vor den wichtigsten archivischen Zugang aus, der zu Beginn einer jeden Archivarbeit durch den Archivar in einem persönlichen Gespräch mit dem Benutzer zu klären ist.

Neue Anforderungen ergeben sich jedoch spätestens dann, wenn man die Findmittel zu den Archivbeständen im Internet zugänglich machen will. Eine individuelle Beratung ist nicht mehr möglich, die archivische

¹ Hubert Höing: ADV-gestütztes Informations- und Dokumentationssystem für Archive (AIDA). Ein vorläufiger Bericht aus dem Bereich der niedersächsischen Archivverwaltung. In: Der Archivar, Jg. 36, 1983, Sp. 149 ff.; Bernd Kappelhoff: Die Nutzung der Datenverarbeitung in den niedersächsischen Staatsarchiven, insbesondere für die Verzeichnung und Erschließung moderner Archivalien. In: Der Archivar Jg. 39, 1986, Sp. 35 ff.; Manfred von Boetticher: Zur neueren Entwicklung der EDV in den niedersächsischen Staatsarchiven. Möglichkeiten des Datentransfers bei AIDA, dem "ADV-gestützten Informations- und Dokumentationssystem für Archive". In: Der Archivar Jg. 46, 1993, Sp. 423 ff.

Tektonik muss hier für sich selbst sprechen. Aus diesem Grund bemüht sich die niedersächsische Archivverwaltung zur Zeit als Ergänzung zur bisherigen edv-mäßigen Erschließung der Bestände um den Aufbau einer Metadatenklassifikation - letztlich einer digitalisierten Beständeübersicht, die sich automatisch aus den jeweils erfassten Beständen aufbaut und die für eine Präsentation im Internet geeignet ist.

Dabei zeigen sich drei zentrale Schwierigkeiten:

- Die über Jahrzehnte gewachsenen Bestandssiglen der Archivbestände weisen eine eher zufällige Reihenfolge auf und entziehen sich weitgehend einem systematischen Zugang.
- Die häufig wechselnden Kompetenzen der verschiedenen Behörden und Dienststellen machen in Verbindung mit der Dimension Zeit Suchstrategien erforderlich, die über den bloßen Behördennamen hinausgehen.
- Die räumliche Zuständigkeit einer Behörde, ihr geographischer Bezug, muss in Verbindung mit der Dimension Zeit bei jeder Suche deutlich gemacht werden.

Im weiteren soll gezeigt werden, welcher konzeptionelle Lösungsweg in Niedersachsen zur Erfüllung dieser Bedingungen eingeschlagen wurde:

Ausgangspunkt der Metaklassifikation stellt eine logische Archivtektonik dar, die es erlaubt, sich von gewachsenen Bestandsbezeichnungen in der notwendigen Weise zu lösen. Ein systematisch aufgebauter Strukturbaum gibt eine Übersicht über die niedersächsischen Staatsarchive und den Aufbau der jeweiligen Bestände (vgl. Abbildung).

Daneben sind in einem weiteren Feld thesaurusmäßig die wichtigsten Kompetenzen sämtlicher Provenienzbildner in alphabetischer Reihenfolge aufgeführt und wie mit einer Suchmaschine - ergänzt um eine zeitliche Festlegung - abzufragen. Als Suchergebnis werden die zugehörigen Be-

standsnamen angezeigt, durch deren Anklicken die jeweilige Stellung des entsprechenden Bestandes im Strukturbaum verdeutlicht werden kann.

Ist ein Bestandsname im Strukturbaum aktualisiert, lässt sich ein Schalter öffnen, der es ermöglicht,

- neben dem Strukturbaum sämtliche einschlägigen Kompetenzen des Bestandsbildners sichtbar zu machen oder
- am unteren Bildschirm weitergehende Informationen zum Bestand zu erhalten oder
- in der Mitte der Bildschirms einen geographischen Bezug (vgl. Abbildung) sichtbar zu machen.

Kartographische Darstellung



Erklärungen zum Bestand

Dies ist der zentrale Aspekt der derzeitigen Planungen: die kartographische Darstellung im archivischen Internetauftritt. Für dieses Vorhaben kann die Niedersächsische Archivverwaltung auf mehrjährige Vorarbeiten der Niedersächsischen Vermessungs- und Katasterverwaltung zurückgreifen, die zunächst bei der Erstellung einer Übersicht über Umweltdaten und Umweltprojekte in Niedersachsen zum Einsatz gekommen sind. Dadurch ist für Niedersachsen ein georeferenzierter Metadatenkatalog vorhanden, der auch für andere Anwendungen zur Verfügung steht, soweit ein Raumbezug hergestellt werden soll.²

Was heißt das? Die Abbildung der Oberfläche des Landes Niedersachsen, die zunächst als Rasterdatenkarte im Maßstab 1:500 000 vorlag, wurde in kleineren Ausschnitten im Maßstab 1:5000 digitalisiert. Es liegen nun Vektordaten vor, die gegenüber den Rasterdaten den Vorteil haben, dass Teile der Karte jederzeit ausgetauscht werden können, da die Informationen objektweise gespeichert werden. Der niedersächsische Landesbetrieb "Landesvermessung + Geobasisinformation" (LGN) verfügt damit über einen auf die jeweiligen Längen- und Breitengrade des Landes, d. h. die Koordinaten, bezogenen Katalog. Aus dieser digitalisierten Grundkarte mit einer Genauigkeit von 3-5 m können beliebige Kartenausschnitte gezogen werden: viereckige bereits durch die Festlegung von zwei Eckpunkten, polygone durch Bestimmung des genauen Grenzverlaufs. Innerhalb dieser Auszüge können beliebige Messungen vorgenommen werden und sind sämtliche dort erfasste Ortsnamen abzufragen.³

² Peter Creuzer/Axel Kleinwächter: Der georeferenzierbare Metadatenkatalog des Landes Niedersachsen, in: Nachrichten der Niedersächsischen Vermessungs- und Katasterverwaltung, 49.1999, S. 99-105.

³ Vgl. Sören Wegener: Die Welt dargestellt aus Bits und Bytes. Digitale Landkarten erhalten Einzug in unser tägliches Leben, ebenda, S. 107.

Von seiten der Archivverwaltung werden nun die historische Grenzen, die den jeweiligen Zuständigkeitsbereich einzelner Provenienzbildner beschreiben, ebenfalls als Vektordaten erstellt und wie eine Folie über die digitalisierten Karten gezogen. Auf diese Weise lassen sich mit relativ geringem Aufwand topographisch genaue historische Karten erstellen, die ihrerseits in Teilen verändert und beliebig bearbeitet werden können. Jeder einzelne frühere und heutige Ortsbezug der Ausgangskarte ist auch in ihnen jederzeit zu realisieren.

Jeder im Programm AIDA erfasste Bestand kann nun nach zeitlichen Schnitten mit seinen entsprechend im Vektorformat definierten Grenzen verknüpft werden. Dabei zeigt sich, dass die Zahl der Grenzen, d. h. die Zahl der erforderlichen Polygone, auf Gemeindeebene sehr bald redundant wird. Im Rahmen der ersten Programmierung sollen etwa 150 historische Karten auf diese Weise erstellt werden. Berücksichtigt werden dabei - von den heutigen Verhältnissen ausgehend - zunächst einmal die administrativen Grenzen der Ämter, Kreise und Regierungsbezirke seit 1946, die Archivsprengel seit 1946 sowie die Gerichtsbezirke seit 1946.

Beim Anklicken eines Bestandsnamens oder einer Bestandssigle sind auf diese Weise Karten sichtbar zu machen, die jeweils nach Zeitschnitten unterschiedliche Zuständigkeitsgrenzen des betreffenden Provenienzbildners anzeigen. Im Rahmen der Metaklassifikation von AIDA sind historisch-geographische Zusammenhänge zu verdeutlichen. Umgekehrt ist es aber auch möglich, die Suche nach einem oder mehreren einschlägigen Archivbeständen im Kartenfeld beginnen zu lassen. Ein Schalter ermöglicht das Festlegen von Zeitschnitten, ein weiterer Schalter das Festlegen von territorialen Zuordnungen oder Ortsbetreffen. Durch ein Wahlfeld ist es möglich, die verschiedenen Zeitfolien beliebig zu kombinieren. Eine gleichbleibende farbige Kennzeichnung soll dabei eine optimale Vergleichbarkeit unterstützen. Durch Anklicken einer bestimmten Karte führt

der Weg jederzeit wahlweise zu den Bestandsnamen im Feld "Kompetenz", zu der entsprechenden Bestandssigle bzw. zur Stellung des Bestandes im Strukturbaum.

Ein Zugriff auf den einzelnen Archivbestand ist damit also grundsätzlich stets möglich über:

- den Behörden-Strukturbaum oder
- die Kompetenzkartei oder
- die Bestandssigle oder
- den geographischen Zugang.

Für künftige Recherchen ist durch den Metadatenkatalog für jede Akte eine Verknüpfbarkeit mit dem gegebenen Geo-Bezug des entsprechenden Bestandes gegeben. Auch ohne Visualisierung der entsprechenden Karten besteht die Suchmöglichkeit nach einzelnen Orten, deren Koordinaten automatisch durch die georeferenzierten Karten erfasst sind - unabhängig davon, ob die betreffenden Orte bei der Verzeichnung genannt wurden oder nicht. Auf längere Sicht würde dies auch bei einer unstrukturierten Recherche die Möglichkeit der Realisierung spezifischer Archiv-Zusammenhänge bedeuten, selbst wenn das Gesuchte im Aktentitel nicht erwähnt wird. Welche Konsequenzen sich daraus für eine künftige Verzeichnung ergeben, bleibt abzuwarten.

Frank M. Bischoff: Das Projekt V.E.R.A. in Nordrhein-Westfalen - Nutzung der Internettechnologie für die Erschließung und archivübergreifende Verwaltung der Bestände

Im Dezember 2000 hat die Archivverwaltung des Landes Nordrhein-Westfalen den Auftrag zur Erstellung eines *Verwaltungs-, Erschließungs- und Recherchesystems für Archive* (V.E.R.A.) erteilt. Das zugrundeliegende Fachkonzept ist im Internet-Informationssystem der Archive in Nordrhein-Westfalen veröffentlicht.¹

Das Fachkonzept zu V.E.R.A. wurde in Kooperation von den vier staatlichen Archiven Nordrhein-Westfalens erstellt. Eine Reihe von technischen Vorgaben hat das Landesamt für Datenverarbeitung und Statistik beigesteuert. Die Durchführung des Projekts erfolgt unter der Federführung des IT-Referats im Ministerium für Städtebau und Wohnen, Kultur und Sport des Landes Nordrhein-Westfalen.

Mit der Realisierung von V.E.R.A. ist die Startext Unternehmensberatung GmbH in Bonn beauftragt. Das Unternehmen hat im Bereich von Archiv- und Dokumentationssoftware einschlägige Erfahrungen und u. a. das Dokumentationssystem HiDA für das Bildarchiv Foto Marburg und Mido-sa-Online für die Marburger Archivschule entwickelt.

Hintergrund und Zielsetzung

Seit 1998 werden die nordrhein-westfälischen Staats- und Personenstandsarchive flächendeckend mit moderner Informationstechnologie ausgestattet. Die Häuser und ihre Außenstellen verfügen über Glasfaser-

¹ Vgl. <<http://www.archive.nrw.de/dok/vera/>>.

netze, zentrale Daten- und Kommunikationsserver sowie eine geeignete Arbeitsplatzausstattung, u. a. mit Microsoft-Office-Produkten. Alle Archive sind an das Landesverwaltungsnetz angebunden und können somit auf E-Mail- und Internetdienste zugreifen. Daneben ist von jedem Arbeitsplatz aus das Empfangen und Versenden von Faxen möglich.

Bis Ende des Jahres 2002 sollen den gegenwärtigen Planungen zufolge rund 250 IT-Arbeitsplätze bereitstehen. Damit können die DV-technische Infrastruktur und die Arbeitsplatzausstattung der staatlichen Archive in Nordrhein-Westfalen als zeitgemäß bezeichnet werden.

Erhebliche Defizite bestehen allerdings im Bereich archivischer Fachanwendungen. Gegenwärtig werden in den vier Archiven Erschließungsprogramme genutzt, bei denen es sich überwiegend um Eigenentwicklungen handelt. Jenseits von Verzeichnungsprogrammen bestehen nur in einem Haus weitere Fachanwendungen; auch hier handelt es sich um Eigenentwicklungen. Lediglich im Bereich der Bibliotheksgutverwaltung² und des Haushalts-, Kassen- und Rechnungswesens³ sind marktgängige bzw. landesweit eingeführte Produkte im Einsatz.

Bereits zu Beginn der Ausstattungsinitiative haben die Archive den Auftrag erhalten, einen Anforderungskatalog für eine geeignete Fachsoftware aufzustellen und ggf. geeignete Produkte zur Beschaffung vorzuschlagen. Damit stellte sich zunächst die Frage, welche Arbeitsbereiche mit einer archivischen Fachsoftware unterstützt werden sollten.

Zumeist werden archivische Fachanwendungen mit Erschließungsprogrammen für Archivalien und Bestände gleichgesetzt. Erschließung ist ein wichtiges Arbeitsfeld der Archive und zugleich Voraussetzung für die

² Bibliotheca 2000 der Fa. B.O.N.D.

³ HKR-Verfahren der NRW Finanzverwaltung.

Benutzung von Archivalien. Ein gutes Erschließungsprogramm mit geeigneter Recherchekomponente für den Lesesaal erleichtert Archivaren und Benutzern ihre Arbeit.

Allerdings entfällt auf die Erschließung nur ein Teil der Arbeitsaufwände, die im Rahmen der Aufgaben staatlicher Archive geleistet werden müssen. Wie eine Erhebung in den nordrhein-westfälischen Staats- und Personenstandsarchiven für das Jahr 1999 zeigt, liegt der Anteil der Erschließung gemessen an allen Arbeitsaufwänden bei 15 Prozent (Abb. 1⁴).

Zugleich ist die Erschließung in Archiven nur ein Bereich von vielen, in denen eine DV-unterstützte Aufgabenerledigung denkbar und sinnvoll ist. Andere Aufgabenfelder bieten sich gleichermaßen an. Im Rahmen der Benutzung müssen Archivalienbestellungen geprüft, Archivalien ausgehoben und reponiert werden. Fremdausleihen müssen erledigt, Rückgabefristen kontrolliert werden. Reproduktionsaufträge von Benutzern müssen in den Fotowerkstätten bearbeitet und nach Abschluss in Rechnung gestellt werden. Im Interesse einer geordneten Magazinverwaltung müssen Lagerorte und Freiflächen nachgehalten werden. Archivalienzugänge und -abgänge müssen bearbeitet und mit geeigneten Kontrollwerkzeugen erfasst werden. Schließlich müssen auch Restaurierungsaufträge kontrolliert und bearbeitet und jede durchgeführte Restaurierungsmaßnahme in einem Protokoll dokumentiert werden. Daneben führen die nordrhein-westfälischen Staatsarchive Geschäftstagebücher, in denen Geschäftsvorfälle erfasst und nachgehalten werden.

⁴ Die Angaben beruhen auf einer Selbstaufschreibung aller Bediensteten der staatlichen Archive in NRW für das Jahr 1999. In Abb. 1 wurden die Daten für verschiedene Aufgabengruppen zusammengefasst. Zur Gruppe der Verwaltungsaufgaben wurden auch Führungsaufgaben, Rüst- und Wegezeiten sowie die Zeitaufwände für Fort- und Weiterbildungsmaßnahmen hinzugerechnet.

In Fällen, in denen eine Dokumentation im weitesten Sinne gepflegt werden muss, bieten sich Datenbank-gestützte Verfahren an. Eine sinnvolle DV-Unterstützung ist darüber hinaus auch in Bereichen vorstellbar, die sich durch weitgehend strukturierte Arbeitsprozesse auszeichnen, etwa Archivalienaushebung und -repositur, Fotoauftragsabwicklung usw. Da für die zuvor genannten Tätigkeitsfelder im Rahmen der Berichtspflicht der Archive jeweils am Jahresende zusammenfassende Statistiken erstellt werden müssen, würde eine DV-Unterstützung nicht allein einen besseren Service und eine verbesserte Geschäftsprozesskontrolle, sondern auch Rationalisierungseffekte mit sich bringen.

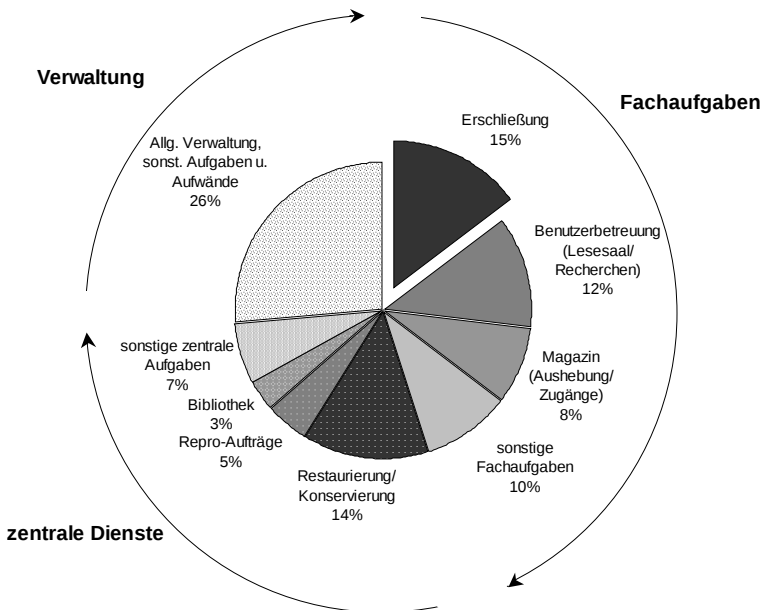


Abb. 1: Verteilung der Arbeitsaufwände in allen nordrhein-westfälischen Staats- und Personenstandsarchiven 1999

Vor dem Hintergrund dieser Überlegungen wurde ein Anforderungsprofil für eine archivische Fachanwendung definiert, das vier Schwerpunkte umfasst:

1. Bestände- und Archivalienerschließung und –dokumentation,
2. Archivgutverwaltung,
3. Benutzungs-, Auftrags-, Kundenverwaltung sowie Recherchewerkzeuge
4. Geschäftstagebuch und Registratur.

Eine bis zum Ende des Jahres 1999 durchgeführte Marktsondierung ergab, dass zwar eine Reihe von Einzellösungen existieren, marktgängige Systeme, die für ein umgrenztes Aufgabengebiet, wie etwa für die Erschließung oder für die Archivgutverwaltung, akzeptable Anwendungen bereitstellen. Allerdings konnte kein System ausgemacht werden, das auf dem aktuellen Stand der Technik das gesamte Anforderungsprofil erfüllt.

Eine Zusammenstellung verschiedener, jeweils für Teilbereiche geeigneter Systeme, schien wenig erfolgversprechend. Die im Anforderungsprofil definierten Schwerpunkte greifen an vielen Stellen ineinander. Dieselben Datenbestände werden in verschiedenen Teilbereichen genutzt. So bestehen etwa Schnittstellen zwischen Erschließungsinformationen, Magazinverwaltung und Restaurierungsdokumentation hinsichtlich der Archivalien oder zwischen Akzessionsverwaltung, Ausleihverwaltung, Fotoauftragsabwicklung und Geschäftstagebuch hinsichtlich der Kunden (Benutzer, Behörden). Eine Entwicklung von Schnittstellen für Systeme verschiedener Hersteller hätte sich kaum erfolgreich zu akzeptablen Kosten realisieren lassen.

Insofern konnte die Entscheidung nur zugunsten einer Neuentwicklung einer umfassenden archivischen Fachanwendung fallen.

Leistungsumfang

Modularer Aufbau

V.E.R.A. soll einen modularen Aufbau erhalten, so dass über die Nutzung einer Teilanwendung – sofern sie nicht für die Funktionsfähigkeit des Gesamtsystems unabdingbar ist – im Einzelfall durch das jeweilige Archiv entschieden werden kann.

Das komplette Leistungsspektrum von V.E.R.A. ist in vier Module unterteilt, die sich ihrerseits wieder in Teilanwendungen gliedern. Modul 1 umfasst im Wesentlichen die dokumentarischen Teile zu Beständen und Archivalien. Im Einzelnen handelt es sich um

- die Verwaltung der Archive und ihrer Bestände,
- die Erfassung von Registraturbildnerinformationen sowie
- die Verzeichnung, differenziert nach verschiedenen Archivaliengattungen.

In Modul 2 ist die Archivgutverwaltung zusammengefasst, nämlich

- Magazinverwaltung,
- Akzessionsverwaltung und
- Kassandenverwaltung.

Die Notwendigkeit einer Kassandenverwaltung besteht bisher vor allem in einem nordrhein-westfälischen Staatsarchiv, so dass diese Teilanwendung von den anderen Archiven vermutlich nur eingeschränkt genutzt wird.

Das Modul 3 umfasst die Benutzerrecherche sowie die Benutzungs-, Auftrags- und Kundenverwaltung. Hierzu zählen

- die Recherchewerkzeuge,
- eine Kunden-, Bestell- und Ausleihverwaltung,

- die Abwicklung von Fernleihen, also die Verwaltung von an die nordrhein-westfälischen Staatsarchive entliehenen Fremdarchivalien und die Fristenkontrolle bei an andere Archive versendeten Archivalien,
- die Erfassung von Sondergenehmigungen sowie eine Sperrfristen- und Zugangskontrolle,
- die Reproauftragsverwaltung und –abwicklung bis zur Rechnungslegung,
- Restaurierungsaufträge und –dokumentation sowie
- Magazin- und Akzessionsaufträge.

Im Modul 3 sind eine Reihe sog. virtueller Aufgabenbücher enthalten, mit denen Aufgaben erfasst und ihre Abwicklung unterstützt wird. Sie kommen in Bereichen zum Einsatz, in denen einerseits verschiedene Handlungsträger an einem Vorgang beteiligt, andererseits die Arbeitsabläufe klar strukturiert sind. Die virtuellen Arbeitsbücher können von den Beteiligten genutzt werden, um sich einen Überblick über den Bearbeitungsstand eines Vorgangs oder über die eigenen Aufgaben zu verschaffen. Sie sollen auch eine konsequente Arbeitsplanung in den Bereichen Restaurierung, Magazinierung, Archivalienreproduktion und Benutzeranfragen unterstützen.

Das Modul 4 umfasst im wesentlichen Geschäftstagebücher, eine Postausgangskontrolle sowie Akten- und Registraturplan.

Einfluss auf Arbeitsorganisation und Erschließungsstruktur

V.E.R.A. ist weder hinsichtlich der Arbeitsorganisation noch im Hinblick auf die Struktur von Erschließungsinformationen ein neutrales Instrument für die Archive. Sofern eine Vorgangsabwicklung durch virtuelle Aufgabenbücher unterstützt werden soll, werden eindeutige Abläufe und Zuständigkeiten vorausgesetzt. Im Detail können diese Abläufe und Zu-

ständigkeiten dann wiederum von Haus zu Haus abweichen, also individuell festgelegt werden. Entscheidend ist allein, dass ein Regelapparat aufgestellt und in V.E.R.A. definiert wird.

V.E.R.A. enthält auch Vorgaben zur Struktur der Erschließungsebenen (Abb. 2). Die verschiedenen Stufen der Erschließung stehen innerhalb von V.E.R.A. in klar definierten Relationen zueinander, die nicht durchbrochen werden können. Daraus sind allerdings in der Praxis keine Nachteile zu erwarten, da die Strukturen am archivfachlich in Deutschland allgemein üblichen Aufbau von Beständen und Findbüchern orientiert sind. Archiv, Beständetektonik und Bestände, Findbücher, Klassifikation und Archivalienbeschreibungen dürften in den meisten Archiven in Zusammenhängen stehen, die V.E.R.A. durchaus abbilden kann. Eine Ausnahme stellt am ehesten die Erfassung von Zwischenarchivgut dar. Zwischenarchivische Funktionen werden von nordrhein-westfälischen Staatsarchiven bislang nur im Zusammenhang mit einer Kassandenverwaltung bis zum Ablauf von Aufbewahrungsfristen wahrgenommen. Entsprechend ist über die Kassandenverwaltung hinaus in V.E.R.A. auch keine Verwaltungsfunktion für Zwischenarchivgut vorgesehen.

Vorgaben für die Archivalienverzeichnung

Hinsichtlich der Archivalienerschließung wird in V.E.R.A. zwischen generellen und speziellen Verzeichnungsinformationen unterschieden. Übergeordnete bzw. übergreifende Elemente sind eine Findbuchverwaltung, eine Indexverwaltung, eine Sperrfristenverwaltung sowie eine Klassifikationsverwaltung. Auf der Ebene der Archivalienverzeichnung können theoretisch bei allen Archivaliengattungen Serien gebildet oder Einzelsvorgänge einer Archivalieneinheit verzeichnet werden.

Bei der Verzeichnung werden Basisangaben zu einer Archivalie von speziellen Erschließungsinformationen in Abhängigkeit von der Archivalien-

gattung unterschieden. Im Bereich der speziellen Erschließungsinformationen sind derzeit definiert: Sachaktenverzeichnung, Urkundenregistrierung und –edition, Siegelbeschreibung, Erschließung von Karten, Plänen und Plakaten, Bild-, Foto-, Film- und Tonmaterialien sowie Personenstandsüberlieferung. Daneben sind drei Gruppen von Massenakten unterschieden, nämlich personenbezogene Akten, Grundbücher und Grundakten sowie Prozessverfahrensakten.

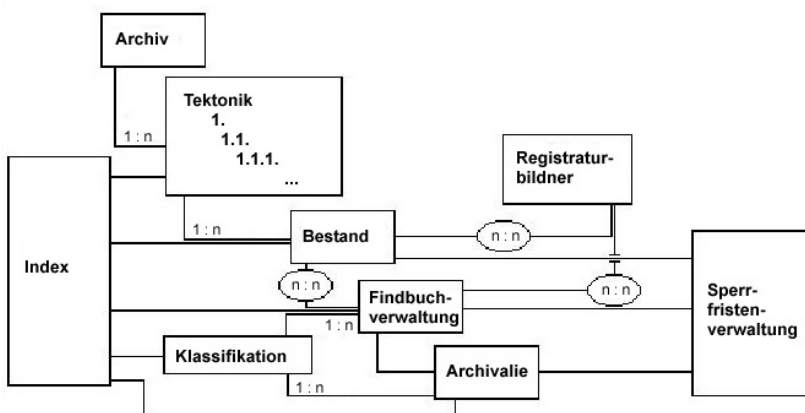


Abb. 2: Struktur von Erschließungsinformationen

Technische Organisation

In allen staatlichen Archiven wird ein Oracle-Datenbankserver installiert, auf dem die lokalen Daten verwaltet, bearbeitet und recherchiert werden. Die Außenstellen sind über ausreichend dimensionierte Standleitungen angeschlossen.

Zugleich wird auf einem Unix-Server im Landesamt für Datenverarbeitung und Statistik eine zentrale Datenbank eingerichtet. Die lokalen Daten werden in regelmäßigen Abständen in die zentrale Datenbank ge-

spiegelt. Somit trägt die zentrale Datenbank einerseits den Charakter einer zusätzlichen Sicherung aller Datenbestände. Andererseits kann ein Benutzer im Archiv A über die zentrale Datenbank auch in den Beständen des Archivs B recherchieren (Abb. 3).

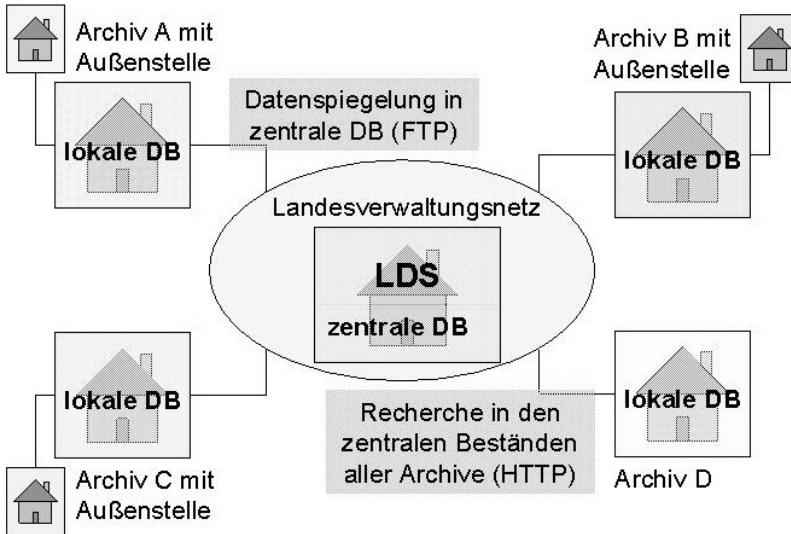


Abb. 3: Technische Organisation

Mit dieser Doppelstrategie wird erreicht, dass die Häuser unabhängig voneinander und unabhängig vom Landesverwaltungsnetz und den zentralen Komponenten im Landesamt für Datenverarbeitung und Statistik arbeitsfähig bleiben. Ein Zusammenbruch des Systems in einem Haus beeinträchtigt nicht die anderen Häuser. Eine Störung im Landesverwaltungsnetz führt zwar zum Verlust der zentralen Recherchemöglichkeiten, tangiert aber nicht die lokalen Arbeiten in den Archiven.

Darüber hinaus sind mit der Einrichtung eines zentralen Servers bereits alle Optionen für eine zukünftige Freischaltung der Bestände im Internet angelegt.

Internettechnologien

Internettechnologien kommen in V.E.R.A. vielfach zum Einsatz. Im Folgenden werden browsergestützte Anwendungen von Internettechnologien im weiteren Sinne unterschieden.

Browsergestützte Anwendungen

Alle Anwendungen in den Lesesälen der Archive werden browsergestützt sein. Diese Entscheidung fußt nicht allein auf einer Orientierung an Standardprotokollen, sondern hier kommen auch Sicherheitsüberlegungen zum Tragen. Der öffentlich zugängliche Teil von V.E.R.A. wird in einem separaten Netz organisiert, über das zwar auch der Zugang zum Internet gewährleistet, ein Zugriff auf die von den Mitarbeitern genutzte Windows-Client-Umgebung aber nicht möglich ist. Mit dieser Entscheidung ist auch die Erwartung verknüpft, dass Benutzer im Umgang mit Internetoberflächen zunehmend vertraut sind und sich so leichter in die verschiedenen Anwendungen in den Lesesälen der staatlichen Archive in NRW einfinden.

Für die Lesesäle sind gegenwärtig folgende browsergestützte Anwendungen vorgesehen:

- Benutzungsanträge,
- Recherchewerkzeuge für lokale Bestände und Findmittel,
- Archivalien- und Bibliotheksgutbestellung,
- Reproduktionsaufträge.

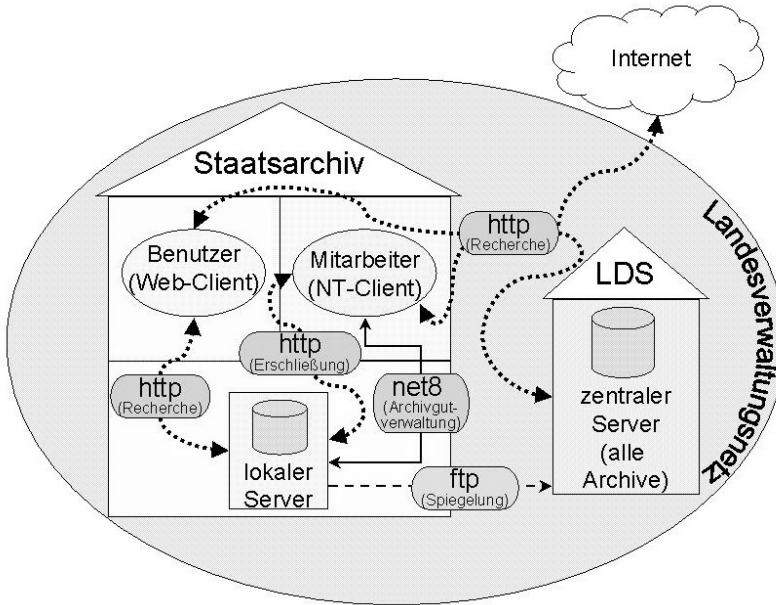


Abb. 4: Datenkommunikation und Kommunikationsprotokolle

Eine weitere browsergestützte Anwendung wurde bereits erwähnt. Die archivübergreifende Recherche in den auf den zentralen Server im LDS gespiegelten Daten aller Archive wird sowohl für Benutzer als auch für Mitarbeiter via http realisiert. Neben den bereits genannten Gründen zur Sicherheit und Benutzerfreundlichkeit spielt hier auch das Datenvolumen eine Rolle, das mit dem Hypertext Transferprotokoll relativ gering gehalten werden kann (Abb. 4).

Internettechnologien im weiteren Sinne

Unter den Internettechnologien im weiteren Sinne sollen hier in erster Linie die Markup Languages verstanden werden. Wesentliche Teile von V.E.R.A. fußen auf XML. Im Folgenden werden Beispiele aus den Berei-

chen der Austauschformate, der Speichertechnik und der virtuellen Aufgabenbücher genannt.

Austauschformate

Während der Arbeiten zum Fachkonzept V.E.R.A. kamen bei den Beteiligten vor allem Bedenken auf, dass Software fast immer bis zu einem gewissen Grad proprietär ist. Ein Plattform- oder Systemwechsel kann in solchen Fällen erhebliche Folgekosten verursachen. Man kann dieser Problematik bei digitalen Systemen nicht grundsätzlich ausweichen. Es herrschte aber rasch Einvernehmen zwischen den Beteiligten darüber, dass Vorkehrungen getroffen werden sollten, um die mit einer Migration verbundenen Gefahren zumindest abzuschwächen. Deshalb verlangt das Fachkonzept eine Reihe von Schnittstellen. Als automatisch generierbare Austauschformate wurden u. a. CSV-Formate, SGML, XML und EAD gefordert. Als Ausgabe- bzw. Darstellungsformate kommen noch RTF und HTML hinzu.

Diese Überlegungen reichen in die Jahre 1998 und 1999 zurück. Die Entwicklung seither hat gezeigt, dass XML als Standard inzwischen weltweit anerkannt wird. Insofern ist innerhalb des V.E.R.A.-Projekts die Bedeutung von XML als Austauschformat deutlich in den Vordergrund getreten, wenngleich auf die anderen Formate nicht verzichtet wird.

Im Hinblick auf EAD sei hier ergänzt, dass eine Übertragung der in V.E.R.A. strukturierten Daten nach EAD erst noch auf die praktische Durchführbarkeit hin getestet werden muss. Bereits jetzt ist aber sicher, dass das im Retrokonversionsprojekt⁵ zu definierende XML-Austauschformat in V.E.R.A. nachgetragen werden soll.

⁵ Vgl. den Projektantrag des Hauptstaatsarchivs in Düsseldorf bei der Deutschen Forschungsgemeinschaft <http://www.archive.nrw.de/dok/retrokonversion01/> sowie die Beiträge von Frank M. Bischoff, Das Projekt Retrokonversion

Speichertechnik

Es wurde bereits erwähnt, dass V.E.R.A. auf Oracle fußt. Oracle ist ein weit verbreitetes, stabiles relationales Datenbanksystem. Datenbanksysteme erlauben es, Informationen in Tabellen, Datensätzen und Feldern zu organisieren, die wiederum relational verknüpft werden können.

Im Rahmen archivischer Erschließung entstehen zum Teil sehr umfangreiche Textmengen. Die Länge eines Aktentitels, eines Enthält-Vermerks, eines Urkundenregests oder einer Findbucheinleitung lässt sich nicht generell auf wenige Zeichen einschränken. In einem Datensatz zu einer Verzeichnungseinheit können mehrere solcher langen Textfelder vorkommen. Unbegrenzt lange Felder mit Text- oder auch Bildinhalten werden in Datenbanksystemen in so genannten Blobs, d. h. Binary large objects, abgespeichert.

Um hier einerseits unabhängig zu sein von etwaigen Beschränkungen von Datenbanksystemen und um andererseits performante Anwendungen zu gewährleisten, wird die Zahl dieser Blobs in V.E.R.A. auf jeweils ein Objekt pro Tabelle reduziert. Das bedeutet, dass alle Informationen zu einer Beschreibungseinheit strukturiert als Blob gespeichert werden, zusätzlich aber noch mit definierten Feldlängen in den üblichen Feldern. Es handelt sich also um eine redundante Datenspeicherung.

Die Strukturierung der Informationen in den Blobs erfolgt mit XML. Das bedeutet, dass alle zu einer Beschreibungseinheit erfassten Angaben in der Datenbank bereits XML-getaggt und ggf. mit Stylesheets definierten Auszeichnungen (fett, kursiv etc.) gespeichert sind.

Nur am Rande sei hier angemerkt, dass auch die Datenkommunikation zwischen Bearbeiteroberfläche, also dem Frontend, und dem Datenbanksystem, also dem Backend, bei den dokumentarischen Daten XML-codiert erfolgt.⁶

Definition der virtuellen Aufgabenbücher

Mit virtuellen Aufgabenbüchern sollen in V.E.R.A. die Arbeitsabläufe in den Archiven unterstützt werden. Hier müssen Transaktionsregeln und Rollen von Akteuren definiert werden. Die Definitionen sollen mit geringem Aufwand geändert und an die Bedürfnisse der verschiedenen Häuser angepasst werden können.

Um die nötige Flexibilität zu erreichen und bei Veränderungen nicht immer eine Nachprogrammierung in Auftrag geben zu müssen, werden hier XML-gestützte Werkzeuge eingesetzt. D. h., dass die Regeln und Rollen in XML-DTD's definiert, die Transaktionen in XML-Dokumenten dargestellt werden und somit durch geschulte Administratoren verändert werden können.

Stand und Perspektive

In 2001 sollen die Module 1 – Dokumentation und Erschließung – und 2 – Archivgutverwaltung – realisiert werden. Gerade wurde das DV-technische Feinkonzept der Fa. Startext abgenommen. Die ersten Programmteile sollen noch im Sommer geliefert werden.

Im Anschluss an die Fertigstellung der Module 1 und 2 sollen in 2002 die Module 3 und 4 beauftragt und realisiert werden. Damit werden die

⁶ Für die Archivgutverwaltung und die virtuellen Aufgabenbücher wird das Oracle-Protokoll *net8* verwendet.

Lesesäle komplett mit digitaler Technik ausgestattet. Außerdem wird die DV-Unterstützung von Arbeitsabläufen umgesetzt.

Die an V.E.R.A. gestellten Erwartungen sind hoch. Zugleich findet das Projekt bei den Kolleginnen und Kollegen in den nordrhein-westfälischen Staats- und Personenstandsarchiven durchweg positive Resonanz.

Es darf aber zweierlei nicht verkannt werden

1. V.E.R.A. stellt hohe Anforderungen an die Archive bzw. ihre Mitarbeiter. Sowohl bei den Systemadministratoren der Staatsarchive als auch bei den Anwendern muss entsprechendes Know-how aufgebaut werden. Vor diesem Hintergrund werden in den Jahren 2001 und 2002 in erheblichem Umfang Schulungen der Mitarbeiter durchgeführt. Die Schulungskonzepte differenzieren zwischen
 - Systemadministratoren,
 - sog. fortgeschrittenen Anwendern, die in die Fachtests eingebunden werden und später für alle Mitarbeiter bei auftretenden Problemen als Ansprechpartner bereitstehen, und
 - Mitarbeitern aus den verschiedenen Arbeitsbereichen, in denen V.E.R.A. zum Einsatz gelangt.
2. V.E.R.A. ist ein mächtiges Instrument, das aber, um seine Vorzüge voll nutzen zu können, mit Inhalten gefüllt werden muss. Bisher liegt nur der geringere Teil der Erschließungsinformationen in einer ohne größeren Aufwand nach V.E.R.A. portierbaren Form vor. Eine wichtige Voraussetzung zur Digitalisierung der bestehenden Findmittel werden die Werkzeuge und Erfahrungen bilden, die im Rahmen des Retrokonversionsprojekts entwickelt bzw. gesammelt werden sollen. Das Beispiel der Bibliotheken lässt allerdings zweifeln, dass allein mit der Verfügbarkeit von Werkzeugen die Archive bereits in die Lage versetzt wären, die flächendeckende Konversion ihrer Findmittel in eine digitale

Umgebung zeitnah aus eigener Kraft zu leisten. Hier werden Archive in Kooperation mit anderen Partnern Konzepte und Strategien für umfassende Konversionskampagnen und deren Finanzierung entwickeln müssen.

Bernward Helfer: Das Hessische Erschließungssystem HADIS 2000

Vorbemerkung¹

HADIS 2000² war der Arbeitstitel für die Novellierung des „Hessischen Archiv-Dokumentations- und Informationssystem“ HADIS (alt), das seit 1987 im Einsatz ist. Wie bereits das ursprüngliche System wird auch HADIS (neu) von den Hessischen Staatsarchiven in Zusammenarbeit mit der Hessischen Zentrale für Datenverarbeitung (HZD)³ entwickelt.

Zur Zielsetzung der Neuentwicklung heißt es im Pflichtenheft: „Die zu unterstützenden Fachaufgaben lassen sich grob mit Verwalten, Verzeichnen und Veröffentlichen umschreiben. Mit der bisherigen Implementierung von HADIS (alt) wurde bereits das Verzeichnen und das Veröffentlichen unterstützt. Neu in HADIS 2000 sind Verwaltungsfunktionen.“.

Dabei sollen:

- die fachlichen Prinzipien, soweit sie sich als tragfähig erwiesen haben, beibehalten,
- die technische Realisierung auf den neuesten Stand gebracht,
- die Altdaten, möglichst verlustfrei, migriert werden.

Die heutige Darstellung bezieht sich auf die Punkte **Verzeichnen** und **Veröffentlichen**.

¹ Bei diesem Text handelt es sich um die überarbeitete Fassung des auf dem 6. archivwissenschaftlichen Kolloquium der Archivschule Marburg gehaltenen Vortrags.

² Im Folgenden unterscheide ich der Einfachkeit halber „HADIS (alt)“ und „HADIS (neu)“

³ Seitens der HZD sind am Projekt beteiligt die Herren Achim Richter und Marcus Nauheimer sowie Frau Ruth Hoffmann.

Aufgabenstellung

Folgende fachliche Prinzipien waren aus HADIS (alt) zu übernehmen:

- Es wird eine Verweis-Datenbank aufgebaut⁴.
- Die Verzeichnung unterschiedlicher Archivalientypen erfolgt über differenzierte Modelle⁵.
- Findbücher werden automatisiert erstellt⁶.
- Es wird ein gemeinsamer Datenpool der drei Hessischen Staatsarchive gebildet⁷.

Neu hinzu kamen die Anforderungen

- Die Betriebskosten der IT-Lösung sollen durch die konsequente Nutzung von Standardkomponenten verringert werden.
- Die Qualität des Datenmaterials soll durch geeignete Strukturierung und Visualisierung, sowie standardisierte Speicher- und Austauschformate verbessert werden.

⁴ D.h. die Datenbank enthält nur Verzeichnungsdaten. Die Verzeichnungsdaten verweisen über die Signatur auf das im Magazin lagernde (Original-) Archivstück. Bei Bedarf kann die Verzeichnung ebenso auch auf eine, wo auch immer vorgehaltene, Speicherung einer Verzeichnungseinheit in digitalisierter Form oder als Bild verweisen. Werden bestimmte Signatur-Richtlinien beachtet, kann ein solcher Verweis automatisiert als Link ausgeprägt sein.

⁵ Die Modelle für Sachakten, Urkunden, Karten und Pläne, Personenbezogenes Schriftgut, Justizschriftgut (Strafsachen, Zivilsachen) wurden nach Hadis (neu) übernommen, nicht zuletzt, da die Alt-Daten migriert werden sollten. Dabei handelt es sich für den Bereich der drei Hess. Staatsarchive zusammen um etwa 1 Mill. Datensätze.

⁶ Dies erfolgte in Hadis (alt) durch Übergabe der Daten von LEDOC an DCF (großrechnergestützte Textverarbeitung). In Hadis (neu) soll hier das Programm Frame-Maker + SGML von Adobe eingesetzt werden.

⁷ Bereits in Hadis (alt) wurde eine archivübergreifende Recherche ermöglicht. Dies soll auch in Hadis (neu) gewährleistet sein.

- Der Service für die Archivbenutzer soll durch vereinfachten und schnelleren Zugang zu den Bestands- und Verzeichnungsinformationen verbessert werden. Dabei soll auch der ortsunabhängige Zugang über das Internet realisiert werden.
- Die Quantität der gespeicherten Informationen soll durch effizienzsteigernde Erfassungsprogramme ansteigen. Insbesondere sollen der Vorgang Datenerfassung und die dabei eingesetzten Verzeichnungswerkzeuge flexibilisiert werden.
- Die Altdaten, insbesondere aus Ledoc (ca. 1 Mill.), aber auch aus Allegro, Access usw. sollen ebenso wie neue Ablieferungslisten oder Scannergebnisse alter Findmittel automatisiert importiert werden.
- Die Zufriedenheit und die Zahl der Anwender soll durch eine moderne und komfortable grafische Benutzerschnittstelle der Programme und durch eine Erweiterung des Funktionsumfangs erhöht werden.

Konzept

- Wir entschieden uns für die Ablösung des bisherigen Dokumentationssystems Ledoc durch ein Programmsystem, das auf der Basis von Windows 95 Standards für das Betriebssystem Windows entwickelt wird.
- Die Datenhaltung erfolgt in einer relationalen Datenbank⁸.

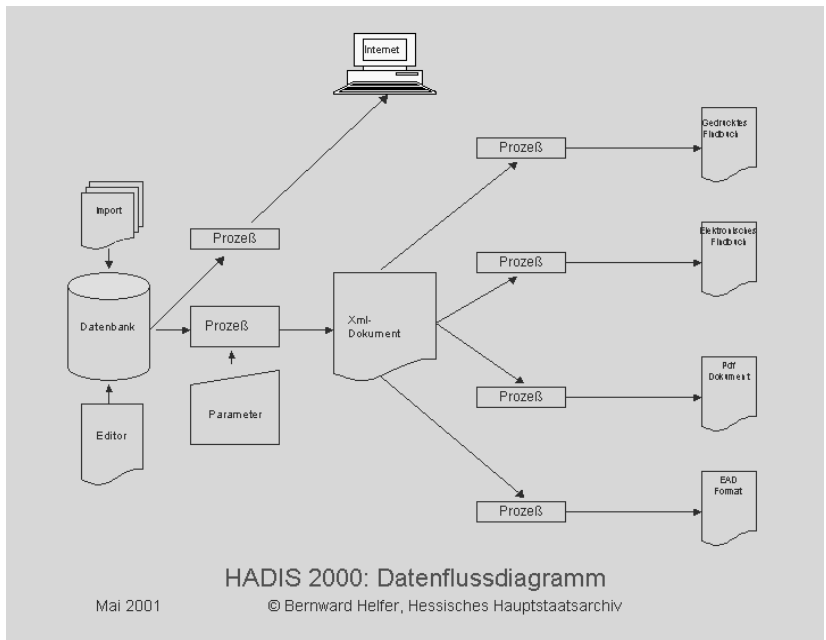
⁸ Zum Einsatz kommt MS SQL-Server. Diese Entscheidung erklärt sich u.a. dadurch, dass zum Zeitpunkt der Entscheidung noch keine nativ XML-fähige Datenbank am Markt war. Das Datenflussdiagramm Abb. 1) zeigt, dass dadurch aufwendige Transformationen notwendig werden, die nicht ohne Auswirkung auf die Performance bleiben. Wie hoch diese Auswirkungen sind und ob die Verwendung einer XML-fähigen Datenbank zu besseren Ergebnissen führen würde, lässt sich z.Z. nicht definitiv sagen.

- Datenspeicherung und Datenausgabe erfolgt über XML-Dokumente.
- Die Datenerfassung erfolgt in „strukturierten Dokumenten“. Sie wird flexibel gestaltet⁹.
- Die Verzeichnung erfolgt nach flexiblen Verzeichnungsschemata¹⁰.
- Die bestandsübergreifende Recherche wird ebenso wie die archivübergreifende Recherche für einen Web-Browser entwickelt¹¹.

⁹ Es hatte sich auch herausgestellt, dass in Abhängigkeit vom zu verzeichnenden Schriftgut mit sehr unterschiedlicher Intensität verzeichnet wurde. Mal wurden sehr ausführliche textuelle Beschreibungen formuliert, mal kam man mit stichwortartigen Angaben aus. Diese Feststellung, verbunden mit der bereits erwähnten Forderung nach schnellem und einfachem Zugang zu den Verzeichnungsdaten (sowohl für die Archivare als auch für die Benutzer), führte uns zu der Entscheidung: es ist eigentlich egal, wie die Daten erfasst werden. Wichtig ist nur, dass sie in der von uns aufgebauten Datenbank enthalten und zugänglich sind.

¹⁰ Die bisherige Arbeit hatte gezeigt, dass die vorhandenen gattungsspezifischen Masken nie ausreichten. Es wurden entweder neue Anwendungen für noch nicht berücksichtigte Überlieferungsarten gefordert (z.B. Tonträger, Amtsbücher) oder zusätzliche Felder in bestehenden Anwendungen für besondere Aktenarten (z.B. Steuerakten, Patientenakten, Gefangenenaakten in der Anwendung Personenbezogenem Schriftgut). Flexible Verzeichnungsmodelle, möglichst unabhängig vom Programmierer, zu entwickeln, war daher erklärtes Ziel.

¹¹ Das bedeutet, dass die Suche sowohl im Internet als auch archivintern mit dem gleichen Programm erfolgt.

**Abb. 1****Datenmodell****Objekte**

HADIS 2000 kennt 3 hierarchisch voneinander abhängige Objekte:

HADIS

- Archive
- - Bestände
- - - Stücke.

Für jedes Objekt wird ein Container gebildet.

Gliederung

Die Objektcontainer können zur Aufnahme der Verzeichnungen mehrstufig gegliedert werden.

Verzeichnung

Eine Objekt-Verzeichnung ist unterteilt in Eigenschaften und Beschreibungen.

Eigenschaften

Jedes Objekt verfügt über (einige wenige) feste Eigenschaften. Da wir Mehrfachverzeichnungen zulassen (siehe unten), werden die Eigenschaften nochmals unterschieden in Objekteigenschaften¹² und Verzeichnungseigenschaften¹³. Darüber hinaus wird ein Objekt durch Beschreibungsmodelle beschrieben.

Beschreibung

Für die redundanzfreie Beschreibung stehen drei Typen zur Verfügung:

- Einzel-Objekt
- Objekt-Serie¹⁴
- Gliederung

¹² z.B. Zugangsangaben, Signatur

¹³ z.B. Laufzeit, Sperrvermerke, Zustandsangaben

¹⁴ Dadurch ist es möglich, bereits innerhalb eines Gliederungspunktes zwecks Vermeidung von Redundanzen eine Datenaggregation vorzunehmen. Ohne diese Möglichkeit bleibt innerhalb des gleichen Gliederungspunktes nur die ständige Wiederholung gleicher Daten, bzw. deren Ersatz durch „dito“ oder ähnliches, also die Redundanz. Eine Aggregation der Daten erfordert dann unausweichlich die Bildung eines zusätzlichen Gliederungspunktes.

Mehrfach-Verzeichnung

Jedem Objekt können beliebig viele Verzeichnungen zugeordnet werden¹⁵.

Beschreibungsmodelle

Das Datenmodell ist so aufgebaut, dass nur die Eigenschaften fest modelliert sind¹⁶.

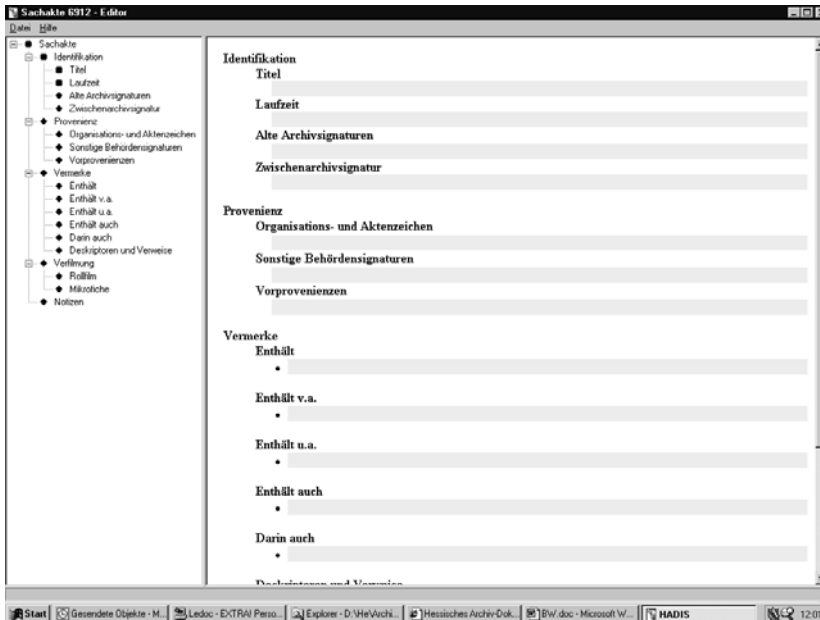
Die Beschreibungen sind dagegen über Modelle oder Schemata frei definierbar. Dies nutzen wir, um für jede Archivaliengattung eigene Modelle zu entwerfen. Diese Modelle müssen dabei keineswegs abschließend definiert sein, sondern können später ergänzt oder variiert werden. Auch kann der Elementumfang eines Modells von Fall zu Fall verändert werden. Beschreibungsmodelle gibt es nicht nur für jedes Objekt sondern auch für die jeweiligen Gliederungspunkte¹⁷.

¹⁵ Dies wird genutzt, um beispielsweise differenzierte Verzeichnungen eines Objekts unterschiedlichen Gliederungspunkten zuordnen zu können, aber auch zur Modifizierung provenienzbezogener Verzeichnungen im Rahmen „Thematischer Inventare“.

¹⁶ D.h. Änderungen sind sehr aufwendig und nur durch einen Entwickler/Programmierer möglich.

¹⁷ Die Gesamtheit der Beschreibungen des Objekts „Bestand“ in einem Archiv ergibt automatisch eine „Beständeübersicht“. Die Gesamtheit der Beschreibungen des Objekts „Archiv“ eine Übersicht der beteiligten Archive (z.B. „Archive in Hessen“). Schließlich ergibt die Gesamtheit der Beschreibungen des Objekts „Stück“ zu einem bestimmten Bestand automatisch ein Findbuch.

Modell Editor



Zur Erstellung der Beschreibungs-Modelle verwenden wir die Beschreibungssprache XML-Data (-reduced).

Dem jeweiligen Datenbankadministrator steht ein Tool zur Verfügung, mit dem er ein neues oder verändertes Modell in die Datenbank einführen kann. Die Prüfung, ob das Modell syntaktisch korrekt ist, übernimmt der Modell-Editor, mit dessen Hilfe die Modelle entworfen werden.

Programm

HADIS (neu) besteht aus einer Vielzahl von Programmkomponenten, die den Zugang zu den Daten, die Erfassung und Pflege der Daten sowie die Publikation der Daten betreffen.

Navigator

Kernstück des Programms ist der Navigator. Ausgehend vom Wurzelknoten HADIS ermöglicht er den (navigierenden) Zugang zu allen Daten. Es handelt sich optisch um eine Adaption des bekannten Explorers. Um die begrenzte Kapazität eines Bildschirms bei der Wiedergabe einer vielstufigen Gliederung nicht zu überfordern, verändern wir beim Wechsel der Objektebene jeweils auch den Wurzelknoten.

Verzeichnen (Datenerfassung)

Editor

Für die (Regel-)Datenerfassung steht ein Editor zur Verfügung. Darin werden die verwandten Elemente (Felder) nicht in einer Maske sondern optisch wie ein einziges Dokument dargestellt. Zunächst navigiert man in die gewünschte Objektebene (Archiv, Bestand, Stück). Dort legt man zunächst ein neues Objekt an und wählt dann ein Beschreibungsmodell aus¹⁸. Jedes Modell enthält Pflicht- und Wahl-Elemente. Der Elementekanon des ausgewählten Modells wird nun als Struktur-Baum im linken Teil eines zweigeteilten Bildschirms dargestellt (vgl. Abbildungen der Beschreibungsmodelle). Die Pflichtelemente sind bereits im rechten Bildteil instanziiert. Aus den Wahlelementen kann der Bearbeiter durch Anklicken die jeweils benötigten Elemente zusätzlich auswählen. Die Feldgröße passt sich automatisch dem Bedarf an und ist praktisch unbegrenzt. Vorgesehen sind die Ausprägungen „Absatz“ und „Liste“. Dies ist für ausführliche textuelle Beschreibungen sicher vorteilhaft, kommt jedoch der Bearbeitung gleichförmiger Massenakten mit wenigen Elementen

¹⁸ Der Vollständigkeit halber sei darauf hingewiesen, dass man statt dessen auch eine weitere Beschreibung zu einem bereits existierenden Objekt erfassen kann.

weniger entgegen. Letztere Arbeiten lassen sich sicher durch spezielle Anwendungen (Access, Excel oder schlicht Wordtabelle) besser erledigen.

Generische Importschnittstelle

Diese Erkenntnis, verbunden einerseits mit der Notwendigkeit, die Altdaten aus Ledoc und z.T. Allegro übernehmen zu müssen und andererseits mit der Absicht, in Zukunft auch Ablieferungslisten in maschinenlesbarer Form oder die Daten eingescannter älterer Findmittel automatisiert zu übernehmen, führten zur Entwicklung einer generischen Importschnittstelle.

Folgende Bedingungen gelten für die Nutzung dieser Schnittstelle:

- die Daten müssen als .csv-Datei (Comma separated Value) vorliegen
- die Daten müssen einem Verzeichnungsmodell zugeordnet werden können. (Fehlt ein solches, kann es eingeführt werden)
- mindestens zwei Angaben müssen als Element „Signatur“ und „Titel“ im Sinne der Darstellung im Navigator definiert werden können.

Beschreibungsmodelle

Sie werden, wie bereits ausgeführt über einen Modell-Editor entwickelt und in die Datenbank eingeführt. Beim Aufruf des Kommandos „Neu“ wird eine entsprechende Auswahl zur Verfügung gestellt. Der Verzeichnungsvorgang braucht also nicht unterbrochen zu werden. Jede Archivierungsart kann jederzeit nahtlos in die Arbeit integriert werden.

Ein Beschreibungsmodell besteht aus einer Vielzahl von Elementen bzw. Elementgruppen. Folgende Beschreibungs-Elemente verdienen eine kurze gesonderte Darstellung.

Laufzeit

Das Element Laufzeit in den Beschreibungsmodellen meint immer Drucklaufzeit, d.h. hier dürfen sowohl alphabetische Zusätze als auch Sonderzeichen bei den Jahresangaben verwandt werden. Der Inhalt dieses Elements wird dann bei allen Wiedergaben angezeigt. Das System bildet aus diesen Angaben zwei weitere Angaben: die Sortierlaufzeit (mit Dialog) und die Suchlaufzeit.

Sortierung

Innerhalb eines Gliederungspunktes (einer Klassifikationsebene) sortiert das Programm automatisch nach der Laufzeit. Bei Serien gilt die Serienlaufzeit. Soll eine von dieser Regel abweichende Sortierung hergestellt werden, muss in einem weiteren Feld ein besonderes Sortiertierkriterium erfasst werden.

Indices und Registern¹⁹

Realisiert ist, für die Weiterverarbeitung durch Frame-Maker, die Eingabe mehrfachgegliederter Indexbegriffe. Vorgesehen (aber nicht abschließend realisiert) ist der Aufbau von benutzerdefinierten Registern in der Datenbank für Recherchezwecke.

Klassifikation

Gliederungspunkte können, wie aus dem Windows-Explorer bekannt, angelegt, verschoben und gelöscht werden. Bei der Benennung ist zu beachten, dass nur die textuelle Bezeichnung anzugeben ist. Die Nummerierung erfolgt automatisch. Dies hat zur Folge, dass auch nach Verschieben eines Gliederungspunktes keine Umnummerierung vorgenommen

¹⁹ Wir unterscheiden zwischen Indices (für gedruckte Findbücher) und Register (für die Suche in der Datenbank).

men werden muss. Die Zuordnung einer Verzeichnung zu einem Gliederungspunkt erfolgt einfach über die Windows-Funktion „Drag und Drop“, d.h. die im rechten Fenster des Navigator enthaltenen Titelaufnahmen werden einem im linken Fenster des Navigator befindlichen Gliederungseintrag zugeordnet. Eine Zuordnung kann beliebig oft verändert werden. Dieser Arbeitsschritt erübrigt sich, wenn die Anlage der Verzeichnung bereits im richtigen Gliederungspunkt erfolgt.

Datenhaltung/-speicherung

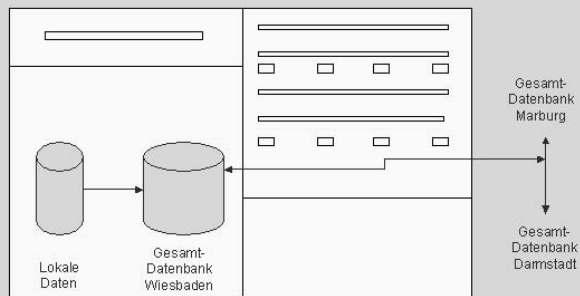
In jedem beteiligten Archiv wird ein eigener Server eingesetzt²⁰. Die (z.Z. drei) Server werden täglich untereinander abgeglichen.

In jedem beteiligten Archiv gibt es zwei SQL-Datenbanken.

- „Vor-Ort-Datenbank“ für die örtlichen Arbeiten (aktuell). Kann von allen genutzt werden, die „dienstlichen“ Zugang haben.
- Gesamtdatenbank mit den Daten der beteiligten Archive für die überörtliche Recherche (tagesaktuell). Diese Datenbank steht auch für die Benutzer im Benutzersaal zur Verfügung.

²⁰ Zunächst war in Analogie zur bisherigen Arbeitsweise in Hadis ein zentraler Datenbank-Server in der HZD in Wiesbaden geplant, auf den die beteiligten (Staats-) Archive über eine Standleitung zugreifen. Entsprechende Tests zeigten sehr schnell, dass die zur Verfügung stehenden 64 KB Leitungen für eine C/S-Applikation (Client/Server) unzureichend sind. Auch eine Verdoppelung der Bandbreite auf 128 KB führte zu kaum nennenswerten Verbesserungen. Es wurden sodann Tests mit Windows-Terminal-Server sowohl von Citrix als auch von Microsoft selbst unternommen. Ergebnisse waren bei ausschließlicher Datenbankbetrieb in Ordnung. Über die gleichen Leitungen ist jedoch auch der Internetzugang der Staatsarchive realisiert. Umfänglichere Arbeiten im Internet (z.B. Down-Loads) führten bei der Datenbankankwendung zu nicht akzeptablen Verzögerungen. Dies alles erbrachte schließlich die Entscheidung für eine dezentrale Lösung.

Muster eines Datenbankservers

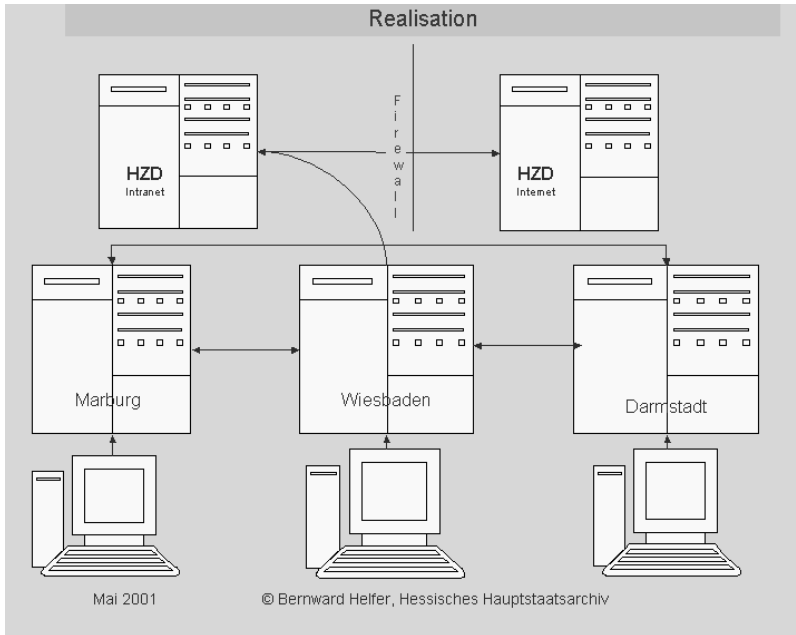


Mai 2001

© Bernward Helfer, Hessisches Hauptstaatsarchiv

Abb. 2

Von der Wiesbadener Version der Gesamt-Datenbank wird der Datenteil, der für die Veröffentlichung im Internet vorgesehen ist, auf einen Internetserver bei der HZD kopiert (tagesaktuell).

**Abb. 3**

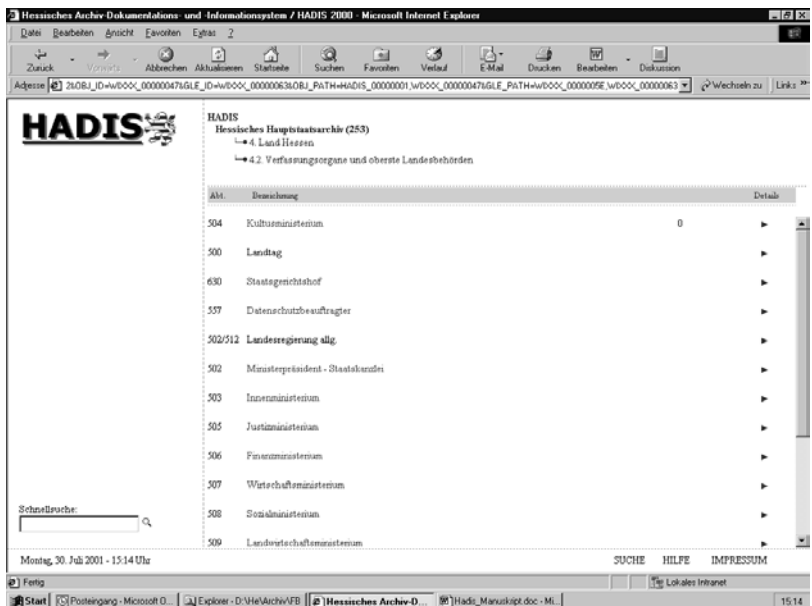
Veröffentlichen

Darstellung im Internet

Ziel dieser Komponente ist es, dem externen Benutzer möglichst umfassende und aktuelle Informationen über das Archiv, seine Bestände und das darin enthaltene Schriftgut zur Verfügung zu stellen. Der Zugang ist datenbankbasiert und erfolgt auch hier zunächst navigierend. Für jedes Objekt stehen alle Verzeichnungen²¹ tagesaktuell zur Verfügung. Durch

²¹ Ausgenommen von der Veröffentlichung im Internet bleiben solche Verzeichnungen, deren Veröffentlichung aus rechtlichen Gründen unterbleiben muss. Hierbei handelt es sich überwiegend um personenschutzrechtliche Gründe, es kom-

die Verwendung eines Web-Browsers als Benutzerschnittstelle erübrigt sich die Entwicklung eines eigenen „Client“²² und die notwendigen Arbeiten „reduzieren“ sich auf die Entwicklung der Server-Erweiterungen.



Die Stückverzeichnisse werden in der in Findbüchern üblichen Form dargestellt. Somit stehen die Verzeichnungsdaten aller Bestände tagesaktuell in Findbuchform zur Verfügung, ohne dass dafür eine Findbuchdatei erstellt werden muss.

men aber auch einschlägige Bestimmungen in Depositaverträgen in Betracht. Dagegen werden im Falle von Sperrfristen nach dem Hess. Archivgesetz die Verzeichnungsdaten mit einem entsprechenden Sperrvermerk veröffentlicht.

²² Dies bedeutet andererseits aber auch Abhängigkeit von Funktionalität und Leistungsumfang des eingesetzten Produktes.

HADIS
Hessisches Hauptstaatsarchiv (253)
Hessisches Ministerium für Wissenschaft und Kunst (511)
↳ Leistungsebene

Nr.	Titel	Lebenszeit	Details
	Auslauterteilung gegenüber dem parlamentarischen Untersuchungsausschuß zum Neubau der Marburger Universitätsbibliothek auf den Lahnbergen	1984-1986	►
58	Leistungsbilanz des Ministeriums <i>Buchalt v.a.</i> - Koalitionsvereinbarung der Landesregierung im Bereich der Kultur- und Wissenschaftspolitik, 1986 - Halbesrieden des BMWK, 1989 - Kulturpolitisches Arbeitsprogramm "Perspektive 91" mit statistischen Materialien, 1989	1985-1989	►
	Korrespondenz mit der Staatskanzlei (nach Sachalphabet) <i>Buchalt v.a.</i> - Abkommen zwischen BfD und DfB auf dem Gebiet von Wissenschaft und Technik - Bund-Länder-Vereinbarung zur Beteiligung der Länder in EG-Öffentlichem - Sitzung des deutsch-amerikanischen Beirats Ausschuss in Kloten Eberbach - Deutsch-französisches Hochschulbeirat	1986-1988	►
	Beantwortung von Anfragen der Landtagsabgeordneten durch das parlamentarische Referat des Ministerbüros (nach Namensalphabet)	1987-1990	►

Schnellsuche:

Montag, 30. Juli 2001 - 15:15 Uhr

SUCHE HILFE IMPRESSUM

Fertig

Status: Posteingang - Microsoft O... Explorer - D:\VieArchiv\FB... Hessisches Archiv-D... Haddi_Manuskript.doc - Mi... Lokales Intranet 15:15

Alternativ zur Navigation können auch Recherchen vorgenommen werden. Hierzu stehen eine

- Schnellsuche und eine
- erweiterte oder detaillierte Suche zur Verfügung (noch nicht realisiert).

Das Rechercheergebnis wird zunächst statistisch aufbereitet. Ist es archiv- oder bestandsübergreifend, ist die Aufstellung entsprechend strukturiert.

HADIS
Hessisches Hauptstaatsarchiv (253)

Dieser Anfrage nach Druckbuch entsprachen 117 Treffer

I. Staatsarchive
253 Hessisches Hauptstaatsarchiv (25)

1.2.1. Reichsritterschaft
2 Mittelheinzliche Reichsritterschaft (4)

1.5.2. Nassau-Usingen
140 Nassau-Usingische Regierung (14)
141 Nassau-Usingische Hofkammer (7)
142 Nassau-Usingische Verwaltung des Fürstentums Nassau-Usingen (7)

1.5.4. Nassau-Weilburg
151 Nassau-Weilburgisches Kabinett (4)
154 Nassau-Weilburg
160 Nassau-Weilburgisches Amt Weilburg (1)

2.3. Zentralverwaltung
202 Herzoglich-Nassauische Militärverwaltung (15)

2.3. Zentralverwaltung
215 Herzoglich-Nassauische Rechnungskammer (7)

2.6.3. Landesarchiv und Forsten
255/2 Herzoglich-Nassauischer Oberforstamt Osnabrück (3)

1.6.3. Hessen-Darmstadt

Montag, 30. Juli 2001 - 15:20 Uhr

Suche Hilfe Impressum

Fertig

Start Posteingang - Microsoft O Explorer - D:\Me\Archiv\FB Hessisches Archiv-D... Hads_Manuscript.doc - M...

Auf Bestandsebene wird das Rechercheergebnis in der HADIS üblichen Findbuchform aufbereitet. Die erste Anzeige erfolgt in einem Kurzformat, bestehend aus Signatur, Titel, Laufzeit.

Jeder Verzeichnungsdatensatz enthält ein „Detail“-Symbol, mit dessen Hilfe der komplette Datensatz aufgerufen werden kann.

HADIS
Hessisches Hauptstaatsarchiv (253)
Domkapitel Mainisches Amt Hechheim (105)

253 105 018

Identifikation
Titel
 Zwerch-, Berg- und Talfahrt auf dem Main zu Eddersheim und Flörsheim
Laufzeit
 1700-1802

Provenienz
Vorprovenienzen
 Mainzer Dompropsteiant

Vermerke
Enthält u.a.
 • Auseinandersetzung zwischen Hessen-Darmstadt und dem Mainzer Domkapitel wegen der Flörsheimer Mannüberfahrt in das Territorium der Herrschaft Eppstein, 1700-1712
 • Protest Hessen-Darmstadts wegen der von Eddersheimer Untertanen und Schiffen vertragswidrig angemaßten Mannfahrt mit einem "Bellhorn", 1757
 • Verpachtung der Eddersheimer Berg-, Tal- und Zwerchfahrt auf dem Main, 1716-1802
 • Beschwerde der Schultheiß von Eddersheim wegen der durch das St. Klara-Kloster zu Mainz von Eddersheimer auf Flörsheimer Schiffleute übertragenen Warentransports auf dem Main, 1760-1762
 • Streit wegen der vom Mainzer Dompropsteiant veranlaßten Arreherung eines im Besitz des preußischen Kaufmanns Johannes Jordan aus Kleve befindlichen Schiffes in Eddersheim, 1784-1786

Schnellsuche:

Mittwoch, 1. August 2001 - 16:34 Uhr

SUCHE HILFE IMPRESSUM

Start Posteingang - Microsoft O... Hads_Manuscript.doc - M... Hessisches Archiv-D...

Hier bietet sich die Möglichkeit, den Datensatz auf einen Merkzettel zu schreiben oder den Kontext aufzurufen. Unter Kontext wird der komplette Gliederungspunkt des jeweiligen Bestands verstanden. Diese Funktion ermöglicht es, das Rechercheergebnis zu verlassen und die weiter oben bereits dargestellten Möglichkeiten des Navigierens zu nutzen.

Findbücher

Gedruckte Findbücher

Der Programmteil „Buchmacher“ erzeugt aus der Datenbank ein XML-Dokument, das an das Programm FrameMaker + SGML übergeben wird. FrameMaker + SGML erzeugt daraus ein Textdokument, das je nach Wunsch bzw. Bedarf in unterschiedlichen Formaten (z.B.: .fm, .ps, .pdf) ausgegeben werden kann.

Elektronische Findbücher

Das von „Buchmacher“ erzeugte XML-Dokument kann aber auch mit Hilfe eines Viewers (z.B. MS Internet-Browser 5.0 oder höher) am Bildschirm als eine Art „Elektronisches Findbuch“ genutzt werden. Diese Möglichkeit ist archivintern vielfältig nutzbar. Ob und wenn ja, wie solche Dateien für einen Down-Load-Vorgang im Internet angeboten werden sollen, ist noch nicht geklärt.

Datenaustausch mit anderen Systemen

Eine Datentransformation zu EAD ist grundsätzlich möglich, zur Zeit jedoch weder realisiert noch beabsichtigt.

Sigrun Eckelmann: Die Archivförderung der DFG – Schwerpunkte und Perspektiven

In meinem Vortrag möchte ich zunächst in groben Umrissen die wissenschaftliche Informationsszene skizzieren, um dann die Archive darin zu positionieren. Daraus lassen sich Anforderungen ableiten, die aus unserer Sicht heute an ein Archiv zu richten sind. Gestatten Sie mir also bitte, eingangs ein wenig auszuholen, um dann auf die Archive und abschließend auf das entsprechende Förderangebot der Deutschen Forschungsgemeinschaft zu kommen.

Die wissenschaftliche Informationslandschaft hat sich in den vergangenen Jahrzehnten grundlegend verändert. Zum einen hat sich der Kreis der Informationsanbieter erheblich erweitert. Wissenschaftliche Bibliotheken, Fachinformationszentren sowie die unterschiedlichen Datenbankanbieter sind die wesentlichen Institutionen, die die Wissenschaftler mit der für Lehre und Forschung benötigten Literatur und Information versorgen. Zum anderen hat der Technologiewandel und hier insbesondere die Entwicklung des Web-Netzes sowohl das Angebot als auch die Funktionalität dieser Informationsdienstleister grundlegend verändert. Im Zuge des zunehmenden Einsatzes netzbasierter Verfahren werden heute nicht nur die bibliographischen Informationen weit überwiegend digital übermittelt sondern zunehmend auch die Volltexte, die entweder eigens retrodigitalisiert oder bereits digital erstellt wurden. Die Verlinkung der Zitate mit den darin genannten Texten ist ein weiterer Service, der vor allem von den Fachinformationszentren immer häufiger angeboten wird. Die traditionelle Bibliothek hat sich so zu einer hybriden Bibliothek gewandelt, bei der die Vermittlung von Dienstleistungen im Vordergrund steht. Neben den gedruckten Informationen bietet sie auch die digitalen

an. Darüber hinaus vermittelt sie Informationen aus netzbasierten Informationsquellen, sorgt für die qualitative Prüfung dieser Informationen sowie für deren langfristige Verfügbarkeit. Neben den gedruckten Werken bietet sie also auch jedwede Form digitaler Informationen und Medien an. Die Weiterentwicklung der Sondersammelgebietsbibliotheken zu „Virtuellen Fachbibliotheken“, Bibliotheken also, die für jeweils festgelegte Fachgebiete neben der gedruckten Literaturproduktion vor allem auch die digitalen Informationen umfassend bereitstellen, wird von der DFG mit hohem Mitteleinsatz gefördert. Mittelfristig entsteht so ein digitales Netz von Bibliotheken. Für die Archivare mag der Aufbau des Informationsservers „Frühe Neuzeit“ bei der Bayerischen Staatsbibliothek München von besonderem Interesse sein. Über all diesen Bemühungen steht das allgemeine Ziel, dem Wissenschaftler in einem sog. „one-stop-shop“ die benötigten Informationen einschließlich der Volltexte ohne Medienbruch zu vermitteln.

Mit den technischen Neuerungen hat sich nicht nur der Informations-transfer grundlegend verändert. Die neuen digitalen Informations- und Kommunikationsnetze schaffen die technischen Voraussetzungen für eine zeit- und ortsunabhängige Kooperation der Wissenschaftler auf nationaler wie auch internationaler Ebene. Die Kommunikations- und Publikationsverfahren verändern sich sehr rasch und nachhaltig und führen zu völlig neuen Arbeitsformen. Vor diesem Hintergrund entwickeln sich u.a. sog. „themenorientierte Informationsnetze“. Sie greifen hinsichtlich der Inhalte weit über die „nackte“ Information hinaus. Hier rücken Produktion und Rezeption des wissenschaftlichen Textes in einen unmittelbaren Zusammenhang. Diese Informationsnetze, die derzeit in vielen Fachgebieten entstehen, werden in aller Regel von den Wissenschaftlern initiiert. Es sind dies Strukturen, die alle Phasen des wissenschaftlichen Produktionsprozesses widerspiegeln:

die *Informationsbeschaffung* – sie bieten bibliographische Daten an und vor allem auch Sammlungen einschlägiger Primärdaten, darüber hinaus Links zu inhaltlich verwandten Institutionen, Veranstaltungen, Forschungsprojekten etc.

die *Kommunikation* – sie bieten Chatlisten, maillisten bis hin zu preprint-Servern an, auf denen die vorläufige Publikation abgelegt wird und bieten damit die Instrumente an, die den wissenschaftlichen kooperativen Produktionsprozess aktiv begleiten,

die *Publikation* – sie bieten die abgeschlossene Publikation an, indem Server eingerichtet werden, auf denen qualifizierte und autorisierte Arbeiten publiziert werden.

Diese Netze sind grundsätzlich auf einen engeren Themenkomplex bezogen, z.B. auf historische Forschungen zum Rhein-Maas-Raum, auf glykowiwissenschaftliche oder neurowissenschaftliche Fragestellungen konzentriert, um nur ein paar Beispiele zu nennen. Oft entstehen sie in Anlehnung an kooperative Forschungsprojekte, sind somit Ausdruck eines neuen kooperativen Arbeitsstils, der in aller Regel international angelegt ist. Sie werden inhaltlich von den Wissenschaftlern entwickelt und fortgeführt, sie werden technisch-organisatorisch in der Regel bei der örtlich zuständigen Bibliothek angebunden.

Insbesondere an dem letztgenannten Beispiel der themenorientierten Informationsnetze wird deutlich, dass das Zusammenspiel zwischen Bibliothek – Nutzer heute sehr viel enger zusammengedrückt, sehr viel enger miteinander verwoben ist, als es früher der Fall war. Daran wird auch deutlich, dass sich die Funktionalität der Informationsanbieter grundlegend verändert hat: ging der Benutzer früher in die Bibliothek, um sich die erforderliche Literatur abzuholen, so werden ihm heute Dienstleistungen an seinem persönlichen Arbeitsplatz angeboten, werden ihm Informationen auf seinen persönlichen Bedarf hin zugeschnitten zusammen-

gestellt. Die Bibliotheken sind nicht mehr nur Bewahrer (gedruckter) Schätze, sie wandeln sich zu Serviceeinrichtungen, die eine breite Palette von Literatur und Informationen anbieten.

Wo stehen hier nun die Archive?

Die öffentlich-rechtlichen Archive leiten ihre Aufgaben aus gesetzlichen Vorgaben ab. Zu ihren zentralen Verpflichtungen zählt, die Schriftstücke der Institutionen ihres Sprengels zu sammeln, zu erschließen und sie den Nutzern zur Verfügung zu stellen. Im Hinblick auf die Wissenschaftler versorgen die Archive die Forscher aller historischen Disziplinen mit archivischen Quellen und stellen für diesen Bereich der Wissenschaften ein unverzichtbares Element im Kreis der Informationsanbieter dar, sind also ebenfalls Teil der beschriebenen Arbeitskette von Informationsbeschaffung, Textproduktion und Publikation. Auch die Historiker stützen sich zunehmend bei der Informationsbeschaffung auf das Internet. Zudem stehen auch sie – wie die Bibliotheken – vor der Situation, dass ihnen die Archivalien zunehmend in digitaler Form zur Archivierung angeboten werden.

Die Anforderungen, die sich aus dem verstärkten Einsatz der Netztechnologie für die Archive ergeben, betreffen sowohl den „Quelleninput“, die Übernahme digitaler Dokumente und ihre Erschließung und Archivierung (hier nur zwei Stichworte: Bewertung, langfristige Speicherung digitaler Dokumente) sowie den „Output“. Ich will mich nur mit dem letzteren Problem, dem „Output“ befassen und meine damit die Präsentation der Erschließungsinformationen sowie die Bereitstellung der Quellen für den Wissenschaftler und die daraus resultierende engere Einbindung der Archive in die Informationsszene.

Die Vorstellung des Archivs mit einer eigenen Homepage und die Präsentation seiner Beständeübersichten im Netz ist gang und gäbe, so dass

ich hier nicht näher darauf eingehen möchte. Der Benutzer erhält so die Möglichkeit, sich von seinem Arbeitsplatz aus einen ersten Überblick darüber zu verschaffen, was er in dem betreffenden Archiv finden kann.

Im nächsten Schritt sucht der Nutzer die Findbücher im Netz. Hier sind noch erhebliche Desiderata zu vermelden. Ziel ist daher, so rasch wie möglich die Findbücher im Netz zu präsentieren. Sofern sie bereits digital erstellt wurden, ist der Weg, sie auch ins Netz zu stellen, relativ leicht gangbar. Instrumente für die Erstellung eines Online-Findbuchs liegen bereits vor, so dass zumindest für die aktuell zu erstellenden Nachweise die Netzpräsentation mitbedacht werden kann. Für die vielen Findbücher jedoch, die nur als maschinen- oder handschriftliche Manuskripte vorliegen, ist ein erheblich höherer Aufwand erforderlich. Erste Untersuchungen finden im Rahmen eines Projekts statt, das die DFG fördert. Es gibt noch viele Probleme - technische, organisatorische und finanzielle - die zu überwinden sind. Bei der Umsetzung in die Breite nach Abschluss dieses Projekts werden ganz wesentlich die Unterhaltsträger gefordert sein.

Inwieweit auch einzelne Dokumente ins Netz gestellt werden, ist sicherlich nicht nur unter technischen sondern vor allem auch unter ökonomischen und archivischen Gesichtspunkten zu diskutieren. Nicht alles, was technisch machbar ist, muss auch sinnvoll sein. Gleichwohl wird es in einzelnen Fällen erforderlich sein, die Originalquellen mit den Zitaten in den publizierten Arbeiten zu verknüpfen, so dass der Leser sich vom wissenschaftlichen Text auf die ursprüngliche Quelle durchklicken kann. Es wäre dieses eine Form der interaktiven Dienstleistung der Archive, die den neuen Arbeitsformen und dem Bedarf der Wissenschaftler weit entgegenkäme. Die themenorientierten Informationsnetze, die derzeit im Bereich der Geschichte entstehen, bieten ein Forum für diese Art der archivischen Informationsvermittlung. Des weiteren wären Verlinkungen zu

anderen historischen Spezialbibliotheken oder Forschungsinstituten denkbar, Verlinkungen der Staatsarchive zu jeweils den historischen Universitätsinstituten ihrer Sprengel. Warum nicht – temporär - Quellen ins Netz stellen, die im Zuge großer (kooperativer) Forschungsprojekte stark benötigt werden? Diese Reihe ließe sich ohne Mühen fortsetzen.

Eine weitere wesentliche technische Voraussetzung für die Verbesserung des archivischen Service ist die Verlinkung der Archive auch über die Landes- und Staatsgrenzen hinaus. Die archivübergreifende Arbeit der Wissenschaftler wie auch Archivare würde dadurch erheblich unterstützt. Hierfür sind nationale und internationale Abstimmung, die Verwendung von Standardsoftware, Standardformaten unabdingbar.

All diesen Zielen ist gemeinsam, die archivische Informationen so umfassend wie möglich im Netz anzubieten und die Interaktion Nutzer - Archiv zu unterstützen. Gleichzeitig ist die engere Einbindung der Archive in die digitalen Informations-Infrastrukturen insgesamt erforderlich. Archive sollten Bestandteil dieser vernetzten Strukturen werden, Bestandteil im „One-Stop-Shop“, das dem Wissenschaftler alle erforderlichen Informationen unabhängig von deren institutioneller Herkunft vermittelt. Die Funktion der Archive als einem wesentlichen Element in der oben skizzierten Informationslandschaft in enger Kooperation mit Bibliotheken, Medienzentren, Rechenzentren und der starken Betonung der Nutzerperspektive sollte neben ihrem Verwaltungsbezug stärker in den Vordergrund treten.

Die Deutsche Forschungsgemeinschaft unterstützt die Archive in mehreren Förderprogrammen beim Übergang vom analogen zum hybriden Informationsanbieter. Die wesentlichen seien hier kurz erwähnt.

Mit der Einrichtung des neuen Förderprogramms *Informations-Infrastrukturen für netzbasierte Forschungsk Kooperation und digitale Publikation* reagiert die DFG auf die bereits oben skizzierten veränderten Arbeitsformen und Rahmenbedingungen und strebt die Förderung des Auf-

baus neuer Kommunikations- und Publikationsnetze an. Wesentlich dafür sind leistungsfähige Instrumente und Infrastrukturen. Dieses neue Förderprogramm hat dementsprechend die Entwicklung prototypischer benutzerorientierter Informations-Infrastrukturen z.B. durch die Weiterentwicklung von prototypischen Standards und Schnittstellen zur Unterstützung dieser neuen Arbeitsformen und damit die Intensivierung der Kommunikation und Kooperation deutscher Wissenschaftler mit dem Ausland zum Ziel.

Gegenstand des Programms *Retrodigitalisierung*, das einen stärkeren Bibliotheksbezug aufweist, ist die retrospektive Digitalisierung ausgewählter Bibliotheksbestände sowie deren überregionale Bereitstellung und Nutzung in einer virtuellen "Verteilten Digitalen Forschungsbibliothek".

Ziel des Programms ist es, durch den Einsatz digitaler Technik die Literatur- und Informationsversorgung zu verbessern. Im Vordergrund stehen dabei:

- der Direktzugriff auf für Forschung und Lehre wichtige Texte,
- der Mehrfachzugriff auf vielgenutzte Literatur und Quellen,
- die digitale Bereitstellung schwer zugänglicher Bestände,
- die erweiterte Nutzung bisher nur wenig bekannter Materialien.

Langfristiges Ziel ist es, die möglichst umfassende Bereitstellung der benötigten Information am Arbeitsplatz des Wissenschaftlers mit der Kommunikation, Verarbeitung und Verbreitung wissenschaftlicher Texte nahtlos zu verbinden.

Mit der Förderung der *Internationalen Kooperationen im Förderbereich „Verteilte Digitale Forschungsbibliothek“* soll ein Beitrag zum Aufbau weltweit verteilter Systeme zur Bereitstellung digitaler Forschungsinformation geleistet werden. Dies erfordert insbesondere

- die internationale Zusammenarbeit bei der Aufbereitung digitaler Inhalte
- die kooperative Entwicklung neuer technischer Werkzeuge und Dienstleistungskonzepte
- die internationale Abstimmung von Konzepten zur Bereitstellung und Verwaltung global verteilter Informationssysteme.

Vor diesem Hintergrund unterstützt die DFG die deutschen Teilnehmer internationaler Kooperations-Projekte mit einem oder mehreren ausländischen Partnern.

Es sei an dieser Stelle auf die derzeit lebhaften Diskussionen von EAD (Encoded Archival Description) hingewiesen. Hier bieten sich internationale Kooperationsprojekte an. Die DFG fördert derzeit bereits zwei Vorhaben zum Thema EAD (Bundesarchiv, Archivschule Marburg), die sich vor allem mit den Einsatzmöglichkeiten von EAD in deutschen Archiven befassen.

Das *Archivprogramm* brauche ich hier nicht näher vorzustellen; es ist Ihnen allen bekannt. Es sei lediglich erwähnt, dass der Förderschwerpunkt heute auf der Weiterentwicklung innovativer archivischer Infrastruktur liegt.

Nähere Informationen über unsere Förderangebote sind unter der Adresse www.dfg.de zu finden.

Schlussbemerkung

Die Zeit drängt. Es täuscht der Gedanke, die Quellen – Unikate – liegen in meinem Archiv, der Nutzer muss ja kommen. Nein, er muss nicht. Es gibt evtl. Gegenüberlieferungen, – oder das Ausweichen auf verwandte Themen/Bereiche/Regionen. Kaum einer ist gezwungen, in ein bestimmtes Archiv zu gehen – von persönlichen Belangen abgesehen. Meines

Erachtens greift es zu kurz, sich auf den Gesetzesauftrag zurückzuziehen, denn auch Gesetze sind änderbar. Es müssen sich auch die Archive mittel- und langfristig – neben den Bibliotheken und anderen Institutionen – auf dem Markt der Informationsanbieter bewähren und dafür neue Instrumente entwickeln. Wollen sie wahrgenommen werden, wollen sie die bei ihnen verwahrten Informationen in dem erforderlichen Maße in Diskussionen aufgenommen wissen, müssen sie sich aktiv an diesem digitalen Informationsmarkt – und es ist heute in der Tat ein Markt – beteiligen. Arbeitsteilung bei Projekten, Vereinheitlichung in der Nutzung und die Erstellung der Nachweise erleichtern den Zugriff und damit die Nutzung der Bestände. Die weitere Entwicklung auch der Archive ist in zunehmendem Maße durch Globalisierung gekennzeichnet. Die digitale Bereitstellung von Forschungsmaterialien intensiviert den globalen Zusammenhang von Forschungsprozessen durch Informationsaustausch und direkte Zusammenarbeit über Netze – gleichzeitig arbeiten die Wissenschaftler zunehmend in internationalen Zusammenhängen und suchen dementsprechend Informationsangebote in weltweitem Rahmen. Die Arbeitsformen und Arbeitsorganisation der Wissenschaftler haben sich dramatisch verändert und verlangen daher auch von den Archiven neue Angebote.

Paul Bantzer: Datenbanken für Internetangebote

Internetangebote erfordern technische Infrastrukturen. Dies gilt auch für Archive, die ihre Informationsangebote im WEB bereitstellen wollen.

Der Beitrag beschäftigt sich mit der Eignung bestehender Datenbankarchitekturen für kulturwissenschaftliche Portale. Besonderer Wert wird auf den Einsatz der verbreiteten Technologien für archivische Anwendungen gelegt.

Datenbanktechnologien

Vier Architekturen sind augenfällig:

- Relationale (SQL) Datenbanken
- Statistische Datenbanken, z.B. OLAP Würfel
- Objektorientierte Datenbanken
- Proprietäre Retrievalsysteme

SQL Datenbanksysteme sind universell einsetzbar und besonders geeignet, aktuelle Informationen zeitnah bereitzustellen und Benutzereingaben unmittelbar zu speichern. Banken und Versicherungen, aber auch Portale mit aktuellen Nachrichten stellen ihre Angebote typischerweise auf diese Technologie ab. Für den Internethandel erfolgt jedoch in aller Regel bereits eine Aufteilung zwischen Informationsangebot und Geschäftsabwicklung: Während das Warenangebot oft statisch im Feilesystem liegt, werden Benutzerberechtigung, Warenkorb und Zahlungsabwicklung über ein relationales Datenbanksystem abgewickelt.

Statistische Datenbanken sind darauf spezialisiert, Zahlen in tabellarischer Form bereitzustellen. Dabei kommt es darauf an, Standardtabellen bereits vorzuhalten und besondere Tabellenanforderungen ad hoc berechnen zu können. Sie finden Einsatz in großen staatlichen Portalen wie

der Gesundheitsberichterstattung des Bundes, aber auch innerhalb großer Firmen, die ihr Data Warehouse im Intranet bereitstellen. Die spezialisierten Anbieter werden zunehmend von den Herstellern relationaler Datenbanksysteme aufgekauft, um die Technologien statistischer Datenbanken in relationale Systeme zu integrieren. Im Bereich der kulturwissenschaftlichen Portale werden statistische Datenbanken kaum eingesetzt.

Objektorientierte Datenbanken haben sich am Markt in nur geringem Maße durchgesetzt. Produktive Anwendungen sind handverlesen. Ein Beispiel ist dem Verfasser aus der Finanzwirtschaft bekannt, ein anderes für die Benutzerverwaltung eines großen medizinischen Dokumentationsportals. Aus technischer Sicht ist die Verwendung objektorientierter Datenbanktechnologien überaus reizvoll, jedoch sprechen zwei gewichtige Gründe gegen den Einsatz: Aufgrund mangelnder Präsenz am Markt sind Analytiker und Programmierer nur selten zu finden, die diese Technologie beherrschen. Der zweite Grund ist das fehlende Zutrauen zur Skalierbarkeit objektorientierter Datenbanken: Sind Internetangebote darauf zugeschnitten, in eine Massenanwendung wachsen zu können, muss die Datenbanktechnologie es ermöglichen, immer neue Server einer Serverfarm hinzufügen zu können. Mit solchem Wachstum haben es objektorientierte Datenbanksysteme schwer.

Die vierte Gruppe von Datenbanken besteht aus den **proprietären Retrievalsystemen**, die vorrangig in Bibliotheken, Archiven und Dokumentationseinrichtungen eingesetzt werden. Es handelt sich um Systeme, die darauf spezialisiert sind, Recherchen in Texten oder geographischen Informationssammlungen durchzuführen. Sie sind für bestimmte Einsatzgebiete spezialisiert und verwenden darauf abgestimmte Daten- und Indexstrukturen. Dabei sind die Systeme darauf optimiert, eine Recherche möglichst effizient durchzuführen. Sie unterscheiden sich in drei Unter-

klassen: Systeme für Volltextrecherche, Systeme für strukturierte Dokumentationsinhalte und GIS-Systeme für die Recherche von geographischer Information. Meistens ist der Aufbau der Indices aufwendig, so dass die Aktualisierung zeitversetzt erfolgt.

Die Art der Vernetzung vorliegender Information bestimmt zu einem gewissen Teil die Auswahl geeigneter Technologien. So ist das V.E.R.A. Fachkonzept konsequent relational strukturiert. Dies hat dazu geführt, Erschließung und Bearbeiterrecherche mit der relationalen Datenbank Oracle zu implementieren. Für die Benutzerrecherche werden jedoch proprietäre Datenbanktechnologien eingesetzt, um den Anforderungen gerecht werden zu können.

Technikausrichtung im Archiv

Im Archiv werden zwei große Arbeitsfelder sichtbar, die datenverarbeitungstechnisch unterstützt werden können. Im Fachkonzept V.E.R.A. sind sie, mit ihren einzelnen Aufgaben, aufgeführt:

- Archivalienzugänge und –abgänge
- Verzeichnung und Findmittelerstellung
- Magazinaufträge
- Benutzeranmeldung, -recherchen, -bestellungen und Ausleihen, Aushebung und Reponierung im Magazin
- Reproaufträge
- Restaurierungsaufträge
- Schriftliche Benutzung, Vorfälle im Geschäftstagebuch und Postausgang
- Sondergenehmigungen
- Virtuelle Aufgabenbücher

Die Arbeitsfelder Verzeichnung und Findmittelerstellung gehören zum kulturwissenschaftlichen Bereich. Die Benutzerrecherche nimmt eine Sonderstellung ein. Alle anderen Arbeitsfelder gehören zur Verwaltung eines Archivs.

Der kulturwissenschaftliche Bereich der Verzeichnung und Findmittelerstellung wird mit Einzelplatzanwendungen wie Midosa bedient. Größere Einrichtungen vernetzen die Arbeitsplätze über Fileserver oder verwenden ein Verzeichnungsprogramm, das in der Client / Server Technologie vorliegt und zur Datenhaltung eine relationale oder objektorientierte Datenbank verwendet. Um die erforderliche Benutzerergonomie der Dateneingabe zu erreichen, sind WEB-basierte Verzeichnungsprogramme vorerst nicht zu erwarten.

Die Sonderstellung der Recherche beruht auf den dokumentarischen Anforderungen, die weit über die Funktionalität relationaler Datenbanken hinausgehen und in aller Regel proprietäre Retrievalsysteme einsetzen. Dabei ist die in den relationalen Datenbanksystemen SQL Server und Oracle enthaltene Volltextrecherche nicht ausreichend, wenn es gilt, in der Suchanfrage Laufzeiten zu berücksichtigen oder die Koexistenz von Sach-, Personen-, Geographie- und Mischindices benutzerfreundlich in der Suchmaske darzustellen.

Dagegen liegt es nahe, die verwaltungsorientierten Arbeitsfelder mit der gleichen Technologie zu bedienen, wie dies in Behörden üblich ist. Die Daten werden in SQL- basierten Datenbanken gehalten, während die Bedieneroberflächen entweder als Windows Client einer Client/Server Architektur realisiert werden, oder, dem Trend entsprechend, WEB-basiert.

Recherche in Findmitteln

Die Anforderungen an ein Datenbanksystem für die Recherche in den Findmitteln sind komplex. Die wesentlichen vier Techniken werden allesamt nur ungenügend von SQL Datenbanken bereitgestellt:

- Volltextsuche
- Suche in Begriffskatalogen, die aus den Indices und Formalangaben gebildet sind
- Navigation in Baumstrukturen über Klassifikation und mehrstufige Indices
- Selektion in geographischen Zusammenhängen

Die **Volltextsuche** wird am besten von proprietären Retrievalsystemen übernommen. Wenn die marktführenden SQL-Datenbanken die Volltextrecherche anbieten, so besteht die Implementation auch dort aus proprietären Fremdkörpern, die - gegenüber dem SQL-Standard - weit entfernt von einer Kompatibilität zwischen den System sind.

Die Benutzeranfragen sind von höchst unterschiedlicher Komplexität. Beispiele sind:

- Einwortabfragen, z.B. der Nachname
- Einfache Phrasen, z.B. „Vorname Nachname“
- Intuitive Verwendung von Fuzzy-Logik in Mehrwortabfragen: die Ergebnismenge wird entsprechend einer heuristisch ermittelten Relevanz dargestellt: je mehr Worte der Abfrage in der Zielinformation enthalten sind, je seltener ein Wort im Dokumentenpool vorkommt, je häufiger das Wort in der Zielinformation selbst genannt ist. Solche Kriterien werden in einen Relevanzfaktor umgerechnet, der selbst bei großen Ergebnismengen die „naheliegendsten“ Zielinformationen zuerst anzeigen lässt.

- Verwendung boolescher Logik mit „and“ und „or“, Proximity-Operatoren, Erweiterung von Begriffslisten über Thesauri und/oder Synonymlisten

Das Blättern in **Begriffskatalogen** entlang sortierter Listen stellt an SQL Datenbanken besondere Herausforderungen, sodass statt dessen eigene Kataloge aufgebaut werden. Die Schwierigkeiten liegen in drei Bereichen:

- Unterschiede in Erscheinungs- und Sortierform: die Zahl 23 sortiert als Textstring kleiner als 4. Umlaute und Ligaturen erfordern eine besondere Aufbereitung. Die Auszeichnung von Texten wie Fett und Kursiv dürfen die Sortierfolge nicht verändern.
- Vorwärtsblättern der nächsten 20 Einträge führt zu ungewöhnlich komplexen SQL-Anfragen, wenn (wie üblich) mehrere Indexabschnitte in einer Tabelle zusammengefasst sind
- Rückwärtsblättern erfordert besondere Vorkehrungen der Indices, die technisch oft nicht vorgesehen sind und zu Hilfskonstruktionen z.B. eines zweiten, jedoch invertierten Katalogs führen.

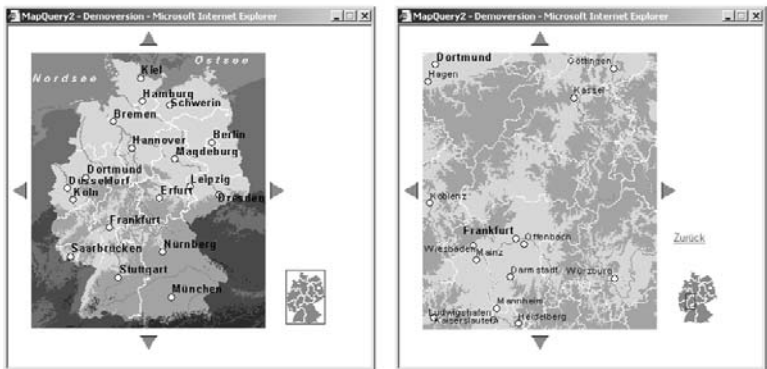
Die **Navigation in Baumstrukturen** der Klassifikation oder mehrstufigen Indices stellt weitere besondere Anforderungen an eine SQL – Datenbank dar. In aller Kürze sei festgehalten, dass relationale Konzepte nur wenig taugen, hierarchische Strukturen abzubilden. Belegen lässt sich dies in der Komplexität und dem Aufwand, der in der Implementation des IBM Content Managers steckt: Das System verbindet die hierarchischen Strukturen mit den relationalen Datenbanksystemen DB2, aber auch Oracle.

Zugehörige Benutzeroberflächen orientieren sich am Explorer für die Navigation im Dateisystem von Arbeitsplatzrechnern. Die Implementation erfolgt als Darstellung von Foldern und Verweisen auf Verzeichnungs-

einheiten. In der Implementation bietet sich für WEB-basierte Oberflächen an, die Folder als XML-strukturierte Einheiten zu generieren und im Browser mit eigenen Scripts zu expandieren.

Die Selektion in **geographischen Zusammenhängen** kann mit GIS Systeme-

Geographische Selektion



men abgebildet werden. Solche Systeme verwenden ebenfalls in aller Regel proprietäre Datenbanken, die außerhalb des SQL Standards liegen.

In der Kosten / Nutzenbetrachtung wird der Zweck entscheidend sein. Die archivfachliche Diskussion führt zu Anforderungen, die mit wirtschaftlichem Aufwand kaum abbildbar sind: Darstellung der Veränderung von politischen Grenzen über die Zeit, und dies in Verbindung mit der Veränderung der Bezeichnung von Gebieten.

Anders sieht es mit intuitiv gestalteten Suchhilfen aus, die dem gelegentlichen Nutzer die Selektion von Beständen ermöglicht. Hier sei ins-

besondere an die publikumsintensiven Personenstandsarchive erinnert, wo der Benutzer an Hand von heute gültigem Kartenmaterial die Ortschaften selektieren möchte, in denen er Aufzeichnungen seiner Vorfahren vermutet. Der dazu erforderliche Aufwand ist vergleichsweise gering: Ergänzung der Verzeichnung der Bestände mit heute gültigen Koordinaten.

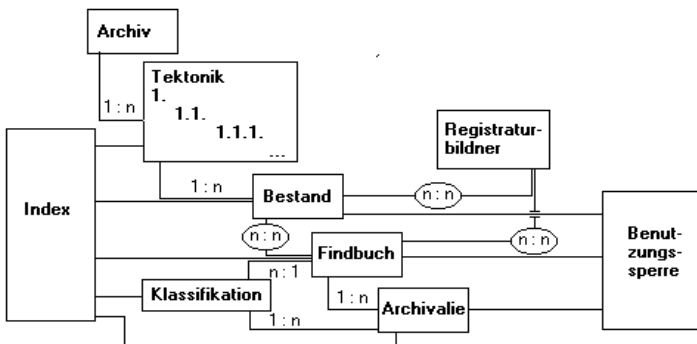
Die dazu erforderliche Recherche über gerasterte Koordinaten ist mit vergleichsweise geringem Aufwand machbar. Angesichts der geforderten Sichtbarkeit und Öffnung der Archive ist so der Nutzen gegenüber dem Aufwand diskutabel.

Unit of Retrieval

Mit der Erstellung von Findmitteln entstehen Informationen, die auf vielen Ebenen suchfähig sind. Genannt seien:

- Abgebende Stellen mit ihren Aufgaben (Registraturbildner)
- Tektonik und Beständeübersichten
- Findmittel mit ihrer Klassifikation

Zur Illustration dient das Schema aus dem V.E.R.A. Fachkonzept:



In der „Unit of Retrieval“ werden alle Informationen zusammengefasst, die als Zielinformation suchfähig sein soll. Die Möglichkeit der findbuchübergreifenden Suche erfordert eine erweiterte Modellierung der Retrievaldatenbank.

So wird es sinnvoll, mit den Informationen der Verzeichnungseinheit alle Informationen zu verknüpfen, die in den darüber liegenden Klassifikationseinträgen enthalten sind. Dabei hat die Suche in der Verzeichnungseinheit ohne ihren Kontext nur wenig Sinn. Deshalb wird die „Unit of Retrieval“ um die Information aus der Klassifikation erweitert.

Die Anreicherung hat jedoch ihre Grenzen: es ginge zu weit, die Information einer Verzeichnung um die Information aus der Einleitung des Findmittels bzw. der inhaltlichen Beschreibung der im Findmittel verzeichneten Bestände zu ergänzen. Hier bietet sich jedoch die mehrstufige Suche an:

Zunächst werden die Findmittel recherchiert, die in die findbuchübergreifende Suche einbezogen werden. Als „Unit of Retrieval“ dient die Beschreibung eines Findbuchs, die um die Beschreibung der zugrundeliegenden Bestände sowie der Aufgaben der abgebenden Stellen erweitert sind.

Diese Anforderungen lassen die Eignung relationaler Datenbanken erneut in Zweifel ziehen: in relationalen Datenbanken ist lediglich der einzelnen Datensatz suchfähig. Auch wenn das Konzept um das Konstrukt des Join erweitert ist, um Informationen mehrerer Tabellen zu suchfähigen Einheiten zusammenzufassen, bleibt die Verwendbarkeit hinter den archivfachlichen Anforderungen zurück: Die Verknüpfung einer Verzeichnungseinheit mit allen hierarchisch darüberliegenden Knoten der Klassifi-

kation zu einem Datensatz sprengt die Grenzen der relationalen Denkweise.

Insgesamt kommen relationale Konzepte nur sehr bedingt an die Anforderungen heran, die eine effiziente Nutzung der archivischen Information erlauben. Im Unterschied zur Nutzung steht die Erschließung: Die Datenverwaltung während dieser Arbeiten erfolgt sinnvoll mittels relationalen Datenbanksystemen, um ausreichende Datensicherheit und die Funktionalität eines Dokumentenmanagement zu gewährleisten.

Abfragemaske

Bei der Abfrage können drei Benutzergruppen unterschieden werden, für die unterschiedliche Benutzeroberflächen sinnvoll sind:

- Archivverwaltung
- Experte
- Gelegentlicher Benutzer

Die Gruppe der **Verwalter** benötigt Masken, die auf ihre Problemstellung zugeschnitten sind. Als Beispiel sei die Recherche des gegenwärtigen Lagerorts einer Archivalie genannt, um den Verbleib nachzuvollziehen: Regal, Reprowerkstatt, Lesesaal oder Restaurator. Diese Masken können mit den Mitteln herkömmlicher SQL - Datenbankprogrammierung erstellt werden.

Die **wissenschaftliche Recherche** erfordert Masken, die einen systematischen, umfassenden und gleichzeitig intuitiven Zugang zu den Suchpfaden der Zielinformation gewährleisten.

Für die beiden Formen der Zielinformationen Bestände und Verzeichnungseinheiten werden die verfügbaren Indexabschnitte mit ihrer Indexierungsart aufgezeigt:

- Volltext

- Mehrwortindexierung (Kataloge, z.B. „Nachname, Vorname“)
- Hierarchische Verzeichnisse (Klassifikation, mehrstufige Indices)

Die mehrstufige Suche wird zwischen den Zielinformationen Bestand und Findbuch unterscheiden. Die Suche wird als Kombination von Navigation, Selektion und Suchanfrage erfolgen.

Der **gelegentlichen Benutzer** ist an der Systematik der Suchpfade weniger interessiert. Er geht intuitiv vor und möchte in seiner Denkweise ziel führend unterstützt werden.

Ein gutes Beispiel stellt die Ebay Maske dar, die im Portal für Privatauktionen verwendet wird.

Auf dieser Maske sind unterschiedliche Techniken vereint:

Die Suche im Volltext der Titel kann mit Checkbox um die Informationen der Beschreibung erweitert werden.

Die Volltextanfrage wird über Drop Down-Combobox mit Operatoren versehen, die der gelegentliche Benutzer nicht beherrschen muss: „Alle Wörter“ (und), „irgendeines der Wörter“ (oder), „exakter Satz“ (adjacent). Ein eigenes Feld ist für Worte vorgesehen, die nicht vorkommen dürfen: „and not“). Auf die relevanzbasierte Suche hat Ebay dagegen verzichtet; denn dieser Mechanismus erscheint für den durchschnittlichen Benutzer zu komplex.

Die Möglichkeit, eine Preisspanne anzugeben, kann für die archivische Suche in die Angabe der Laufzeit übersetzt werden.

Der geographische Standort bietet sich als Verzweigung in die Suche der abgebenden Stelle. Bei Personenstandsarchiven kann hier die aktuelle Karte als Suchhilfe dienen.

Die Kategorien verzweigen in die Klassifikation. Im Fall einer findbuchübergreifenden Suche wird diese Möglichkeit jedoch kaum Anwendung finden.

Die Darstellungsform der Ergebnismenge sowie die Sortierfolge ermöglichen die Spezifikation der Suche auf nur einer Maske.

Die Vielfalt der Parameter auf der Suchmaske ergibt eine gleichermaßen durchdachte wie auch einfache Oberfläche für den gelegentlichen Benutzer. Das ist entscheidend, denn die Akzeptanz des WEB-basierten Zugangs ist der Schlüssel zur Sichtbarmachung und Öffnung der Archive für ein breiteres Publikum.

Gerhard Knorz: Visualisierung von Zusammenhängen - Von der Wissenskarte zur interaktiven graphischen Topic Map

Visualisierung: Notwendigkeit oder Irrweg

Das exponentielle Wachstum der Nutzung des Internets für private und kommerzielle Zwecke hat in bisher kaum gekanntem Ausmaß Überfluss und Mangel schwer trennbar zusammengebracht. "Vor lauter Bäumen den Wald nicht mehr sehen" oder "die Suche nach der Stecknadel im Heuhaufen" beschreiben als Metaphern des Alltags zutreffend die Situation, in der sich erfahrene Nutzer von Informationstechnologie gleichermaßen befinden wie Personen, für die Technik nicht im Vordergrund stehen darf, wenn es um die Erledigung ihrer Aufgaben oder aber auch etwa nur um Unterhaltung geht.

Ein Internetnutzer, dem 10.000 scheinbare Suchergebnisse präsentiert werden, ein Kunde, der potentiell das gesamte Warenangebot eines Versandhandels im Zugriff hat, ein Mitarbeiter in einem Unternehmen, dessen Intranet mehr als 1000 Datenbanken der verschiedensten Abteilungen integriert, reagiert in vielen Fällen hilflos. Sein Problem besteht in einem Überangebot an Daten. Von entscheidenden Fortschritten dahingehend, Orientierung und Übersicht zu schaffen und Zusammenhänge transparent zu machen, wird der Erfolg von E-Commerce mit abhängen. Der Aufstieg des Buzzwords "Wissensmanagement" belegt, dass das angesprochene Problem zusammen mit anderen zu den zentralen gegenwärtigen Herausforderungen gehört.

Das Problem, dass sich eine überbordende Fülle von Daten und Fakten unmittelbar von Menschen schwer beherrschen lässt, hat bereits seit langem systematische Forschungen auf sich gezogen. Grundgedanke ist,

dass das menschliche Gehirn wesentlich besser mit visuellen Eindrücken als mit abstrakten Zahlen oder Fakten umgehen kann [Eib 2000, S. 43 ff], was sich auch an seiner Anatomie ablesen lässt und evolutionsgeschichtlich begründet werden kann [Ror&Lun 1999, S. 790f]. Eine erste Definition von "Visualization in Scientific Computing (ViSc)" geht zurück auf eine Computer Graphic Tagung der NSF im Jahr 1986 [Eib 2000, S. 43]. In der Zwischenzeit hat sich eine differenzierte Begrifflichkeit herausgebildet, die hier vereinfachend auf folgende Feststellung reduziert werden soll [Car et al. 1999]: Computerbasierte Visualisierung bezeichnet die in der Regel interaktive grafische Umsetzung von Daten. Handelt es sich bei den Daten um physische, so spricht man von "scientific visualization", handelt es sich um abstrakte bzw. nicht-physische, so spricht man von „information visualization“.

Der Visualisierung, die die Anforderungen an den mentalen Interpretationsvorgang vom analytischen zum visuellen Teil der Kognition verschiebt, wird ein großes Potential zugetraut. Es ist Teil der Alltagserfahrung, dass die Erkennung von Mustern in Daten (Gruppierungen, Häufungen, Formen, Trends) durch Visualisierung stark vereinfacht (oder erst ermöglicht) wird und dass Zusammenhänge zwischen Daten beziehungsweise Beziehungen zwischen den Daten und ihren Dimensionen durch grafische Präsentationen außerordentlich gut verdeutlicht werden können. [Däß&Pal 1998, S. 41] prognostizieren: „Multiple Informationsräume, eine Kombination aus physisch-raumbezogenen und abstrakt-symbolischen Daten, werden die Benutzerschnittstellen der nächsten Generation von Informationssystemen maßgeblich beeinflussen.“

Andererseits darf die naive Annahme, Visualisierung sei einer textuellen Alternative in jedem Fall vorzuziehen, als überholt angesehen werden. Schon 1983 warnte Tufle vor der Gefahr kontraproduktiver Visualisierung: "Show data variation, not design variation." [Tuf 1983]. Er lieferte

mit seinem Maß des „Daten-Tinte-Verhältnisses“ nicht nur einen konstruktiven einschlägigen Beitrag, sondern prägte auch das eingängige Etikett „chart junk“. Visualisierung hat ihre eigenen konzeptionellen wie auch technischen Probleme und ihre Nutzung als interaktive Benutzungsschnittstelle weitet diese Probleme deutlich aus. „Die Güte einer Visualisierung kann letztendlich nur in extensiven Anwendertests ermittelt werden. Ohne eine solche Güteprüfung ist ihr Einsatz nicht zu rechtfertigen“ [Eib 2000, S. 90]. Oder wie Däßler kritisch formuliert: „The visualization of inherently abstract information is much more difficult and the effect in user interface design less well understood as the role of scientific visualisation“ [Däß 2000, S. 105]

Die gegenwärtige Situation ist also ein Spannungsfeld zwischen Aufbruchstimmung und Ernüchterung, zwischen neuen technischen Möglichkeiten und der Unsicherheit darüber, wie diese Möglichkeiten ausgestaltet werden müssen, damit tatsächlich ein Fortschritt für den Anwender erreicht werden kann. Als ein außerordentlich produktiver Motor für viele aktuelle Entwicklungen hat sich der Erfolg von XML (Extensible Markup Language) herausgestellt. XML zielt unmittelbar auf die Anforderungen an das und im Internet. Es hat die umfang- und facettenreichen Konzepte von SGML (Standard Generalized Markup Language) auf ein Maß reduziert, das es erlaubt, mit angemessenem Aufwand effiziente Werkzeuge zu entwickeln. Damit ist das Konzept von SGML, die Trennung von Inhalt, Struktur und Layout, innerhalb kürzester Zeit in einem offenen und dynamischen Markt populär geworden und hat ganz allgemein auch das universelle Problem des Datenaustauschs für sich vereinnahmt. XML liefert die Infrastruktur für die Kommunikation unterschiedlichster verteilter Anwendungen.

Es lässt sich beobachten, dass unter dem Label „Wissensmanagement“ und unterstützt durch die neuen Möglichkeiten der einfachen Integration

verschiedener Programmkomponenten (sowie durchaus noch weiterer Bedingungen, auf die hier nicht eingegangen werden soll) nunmehr Ideen und Konzepte in neuer Realisierung eine breite Wahrnehmung und Popularität erfahren, die sie zur Zeit ihrer (erstmaligen) wissenschaftlichen Entwicklung auch nicht annähernd erreichen konnten.

Der Beitrag will exemplarisch zeigen, wie zwei Themen durch die gemeinsame Klammer XML einerseits ganz eng zusammenrücken können und andererseits beide für die Zukunft der Weiterentwicklung des Internet und seiner Dienste ein sehr großes Potential aufweisen. Diese Themen sind

- Topic Maps, ein XML-basiertes Austauschformat für semantische Netze und
- SVG, ein XML-basierter Standard für zweidimensionale interaktive graphische Schnittstellen.

2 Beispiele von Visualisierungen

Im Kontext der Herausforderung „Wissensmanagement“ haben so genannte "Wissenskarten" sich als effektives (und gleichzeitig low tech) Werkzeug empfohlen und bewährt. Ein besonders eindrucksvolles Beispiel ihres Potentials ist in [See 1997] zu finden. Wissenskarten visualisieren Wissen über Wissensquellen und Wissensträger. Sie sind Meta-Information zur Erhöhung der Wissenstransparenz. "In its simplest form, a knowledge map is a straightforward directory pointing people who need access to knowledge to places where it can be found" [See 1996].

Sie "vereinigen in sich den Ansatz der Visualisierung von Wissen, das Hypermedia-Konzept mit seinen Navigationspotentialen sowie möglichst die technologiegestützte Gestaltung von Geschäftsprozessen mittels Workflow Systemen, Groupware oder im Intranet. Eine grafische und zugleich strukturierte Darstellungsweise von Wissen ist dabei die Leit-

idee der Kodifizierung." [Noh 2000]. Das Beispiel einer Wissenskarte (die sich gleichzeitig auf den Prozess ihrer Erstellung bezieht) findet sich bei [Kno 1999]. Technisch sind Knowledge Maps vielfach als maussensitive Image Maps als Teil von HTML-Dokumenten ausgeführt (Beispiel unter [Kno 2000]).

Wissenskarten sind i. d. R. Visualisierungen, die das Ergebnis intensiver intellektueller Analyse und einzelfallbezogener Gestaltung darstellen. Themenkarten dagegen (concept map, theme map, tree map, self organizing maps) dagegen entstehen, indem eine Dokumentenkollektion (a priori vorgegeben oder a posteriori als Ergebnis einer Recherche) aus ihrem n-dimensionalen thematischen Raum auf eine zwei- bzw. dreidimensionale Karte heruntergerechnet wird.

Das folgende Beispiel einer Themenkarte ist mit dem Tool Websom erzeugt (<http://websom.hut.fi/websom/>), dessen größte publizierte Anwendung über 6 Mio. Patentdokumente abdeckte [Koh et al. 2000].

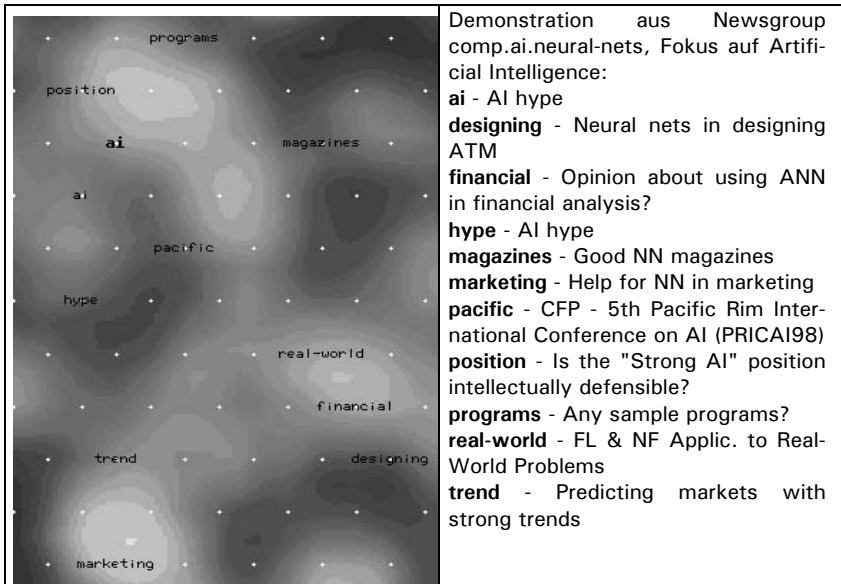


Abb. 1: Mit Websom erzeugte Themenkarte (interaktive Demonstration unter <http://websom.hut.fi/websom/>). Die Themen-Cluster werden automatisch durch ein Schlagwort benannt (rechte Seite). Per Klick kann sich der Nutzer in die Karte und in mehr Details hineinzoomen.

Themenkarten können in naheliegender Weise als Recherchetool verwendet werden, und für das Recherchieren wurden in den letzten Jahren eine Vielzahl von auf Visualisierung beruhenden Ansätzen vorgeschlagen und entwickelt:

- Lyber World, [Hem 1993], Screen Shots unter <http://www.iud.fh-darmstadt.de/Methodik/publ/slide/vvdm1.htm#IMG4>;

- TOSCANA, [Kol et al. 1994] , Screen Shots unter [http://www.mathematik.tu-darmstadt.de/~vogt/](http://www.mathematik.tu-darmstadt.de/~vogt/software/toscana-hardcopy-dt.html) software/toscana-hardcopy-dt.html
- DEViD (Graphik Design und Softwareergonomie integrierende Visualisierung für Document Retrieval Systeme), [Eib 2000];

Weitere Systeme werden im Überblick dargestellt in [Eib 2000] und [Däβ&Pal 1998].

Es wird vielfach diskutiert, ob die Probleme (und die vorgeschlagenen Lösungen) des Wissensmanagements tatsächlich neu sind. Unzweifelhaft neu ist jedenfalls die Tatsache, dass im Forschungsbereich entwickelte einschlägige Lösungsansätze nunmehr tatsächlich erstmals in kommerziell verfügbare und nachgefragte Lösungen umgesetzt werden.

So gibt es seit einiger Zeit für das Problem, Netzwerkstrukturen grafisch-interaktiv zur Exploration bzw. zur Navigation anzubieten, Produkte, über die man sich im Internet eine Meinung bilden kann. Diese Produkte unterscheiden sich zunächst darin, von welchen Arten von Netzwerken/Hierarchien sie ausgehen. Sie differieren auch darin, welche Prinzipien sie heranziehen, um einerseits einen Überblick über einen möglicherweise sehr umfangreichen Strukturkontext zu liefern und andererseits eine handhabbare und übersichtliche Fokussierung zu erreichen. Bekannte Beispiele solcher Produkte sind zu finden unter

- <http://www.visualthesaurus.com>, (Ein assoziatives semantisches) Netzwerk mit getypten Knoten und gewichteten Kanten)
- <http://www.thebrain.com/>, (Ein Browser für Netzwerkstrukturen)
- <http://www.inxight.com/>, (Eine umfangreiche Website mittels "Star Tree" durchwandert; hyperbolische Darstellung)
- <http://www.i-views.de> (Ein Werkzeug für den Aufbau und die Nutzung von semantischen Netzen (K-Infinity))

3.1 Semantische Netze

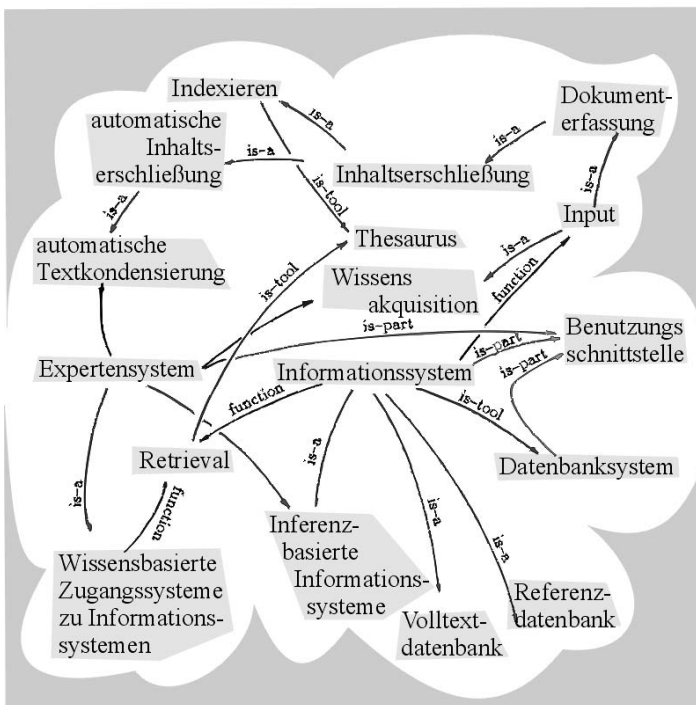


Abb. 2: Ein semantisches Netz mit den Beziehungstypen is-a (Abstraktionsrelation mit Vererbungsmechanismus), is-part (Teil/Ganzes) und anderen (function, is-tool). Die Abbildung ist aus <http://www.k-n-o-r-z.de/publ/slide/fdwrsn1.htm> in vereinfachter Form entnommen.

Semantische Netze, eingeführt in [Qui 66], beherrschten und beeinflussten lange Zeit das Feld der Wissensrepräsentation im Bereich der künstlichen Intelligenz [Bra&Lev 1985]. Mittlerweile verwenden Unternehmen, die Produkte auf der Basis von semantischen Netzen anbieten, auch neue Benennungen wie „Wissensnetze“ oder „Themennetze“. Diese Netze modellieren Wissensausschnitte als Graph aus Knoten und gerichteten Kanten: Knoten zur Darstellung von Objekten (Begriffen) und Eigenschaften sowie gerichteten Kanten (Links), mittels derer die Knoten miteinander in Verbindung gebracht werden.

Inferenzregeln, die beispielsweise die Transitivität einer Relation ausnutzen, können neues Wissen aus „altem“ erschließen. Knoten und Kanten erhalten Benennungen. Semantische Netze eignen sich gut, vernetztes Wissen und kontextabhängige Sachverhalte darzustellen. Der Vorteil der semantischen Netze liegt speziell in seiner Flexibilität und Eingängigkeit.

Genau für diese Vorteile gibt es aber auch eine Kehrseite der Medaille: Die Übersichtlichkeit, die von kleinen graphischen Darstellungen semantischer Netze suggeriert wird, geht mit zunehmender Zahl der einschlägigen Begriffe schnell verloren. Eine Lösung für dieses Problem bieten die in [Min 1975] vorgeschlagenen „Frames“, die man sich als ein gecluster-tes semantisches Netz mit zusätzlichen Strukturelementen vorstellen kann. Frames entsprechen dabei Teilnetzen und repräsentieren im wesentlichen Objekte mit ihren Eigenschaften. Untereinander sind Frames hierarchisch organisiert, wobei untergeordnete Frames (Unterbegriffe) von den übergeordneten (Oberbegriffen) Eigenschaften und Verhalten erben.

Der zweite Nachteil semantischer Netze beruht auf der Leichtigkeit, neue Klassen von Beziehungen zu definieren und zu verwenden, ohne dass deren Definition tatsächlich den Ansprüchen genügt, die formale Inferenzmechanismen voraussetzen. Das bedeutet, dass die natürlichsprach-

lichen Benennungen dem menschlichen Interpreten eine Semantik suggerieren, die tatsächlich in der Struktur des Netzes gar keine Entsprechung hat [Fav 2001, S. 73]. Diese Kritik hat zu den vielverzweigten Arbeiten an Strukturierten Vererbungsnetzen geführt (SVN), deren prominentester Vertreter KL-ONE [Bra&Sch 1985] ist – mit Nachfolgern wie KL-TWO, CLASSIC, BACK und SB-ONE.

Dennoch, gerade weil semantische Netze so flexibel sind, sind sie gewissermaßen durchlässig auch gegenüber abgewandelten oder alternativen Formen der Wissensrepräsentation, so dass sie tatsächlich für eine große Bandbreite von Anwendungen vielversprechend eingesetzt werden können.

3.2 Topic Maps

Das Konzept semantischer Netze scheint inzwischen vor einer Phase erheblich verbreiteter Anwendung zu stehen. Der Grund heißt „Topic Maps“ und stellt den vorläufigen Endpunkt einer Entwicklung dar, die sich in den frühen 90er Jahren unter dem Namen „Topic Navigation Maps“ aus dem absehbaren Bedarf an intelligenten Indizes für Informationsressourcen und aus deren (automatischen) Integration ergeben hat ([Rat&Pep 1999] nach [Hal 2001, S. 8]). Der dann im März 1999 unter der aktuellen Bezeichnung „ISO/IEC FCCD 13250:2000 – Topic Maps“ verabschiedeten Sprachdefinition folgte dann im Dezember 2000 auf der XML-Konferenz in Washington die erste zur Verwendung im WWW angepasste Syntax XTM 1.0 ([XTM 2000]) [Hal 2001, S. 8, 13]. Von einer dynamischen Weiterentwicklung ist auszugehen. Mit Topic Maps sollen Informationsressourcen an eine semantische Modellierung ihrer Inhalte angebunden werden können mit dem Endziel eines „semantischen Web“. Kurzgefaßt läßt sich sagen: Topic Maps sind ein XML-basiertes Austauschformat für semantische Netze.

Topic Maps bestehen zentral aus den so genannten Topics, die einen beliebigen Inhalt (Person, Ort, Ereignis, Begriff) repräsentieren und durch eine ID stets eindeutig identifiziert werden. Topics können in unterschiedliche inhaltliche Zusammenhänge gestellt werden. Dazu gibt es Typen (Types), Instanzen (occurrences) und Beziehungen zwischen Topics (Associations). Diese Ausprägungen werden einem Topic jeweils in Form von Attributen zugewiesen, die ihrerseits selbst wieder Topics darstellen können. Weitere führende Konzepte sind Scopes (Gültigkeitsbereiche) und Public Subjects (allgemein bekannt Themen). Mit Topic Maps können auf den gleichen Informationsbestand, der grundsätzlich nicht Bestandteil der Topic Map ist, viele verschiedene Sichten definiert werden [Hal 2001, S. 9].

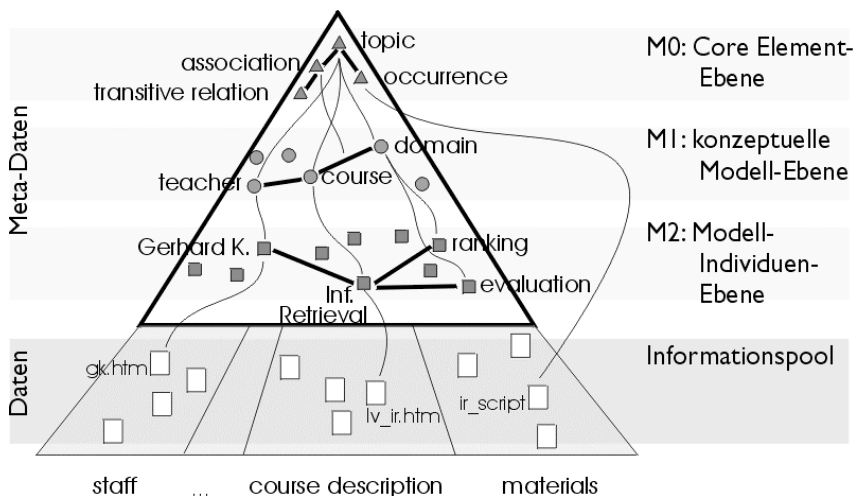


Abb. 3: schematische Darstellung der Modellebenen einer Topic Map. In Anlehnung und Weiterentwicklung an [Hal 2001, S. 33]

Anhand von Abbildung 3 soll die Architektur einer Topic Map-Anwendung erläutert werden. Eine derartige Anwendung für die Verwaltung

von Kompetenzen an einem F&E-Institut für den internen Review-Prozess findet sich in [Hal 2001] detailliert entwickelt und beschrieben. Die abstrakteste Modellebene M0 beschreibt die XTM-Elemente, die verwendet werden, um die konzeptuellen Modelle M1 für den zu Grunde liegenden Informationspool zu definieren. Die Ebene M0 wird zunehmend durch Bezug auf die XTM Core Elements (<http://www.topicmaps.org/xtm/1.0/core.xtm>) abgedeckt werden können. Das konzeptuelle Modell M1 wiederum definiert vergleichbar einem Datenbankschema die Elemente (Objektklassen, Informationstypen, Beziehungstypen, Rollen und Funktionen), die für die Anwendung benötigt werden. Hier wird etwas festgelegt (siehe Abb. 3), dass es *teacher* als *Lehrende* in *Kursen* gibt, um dann auf der Individualebene M2 das Faktum definieren zu können, dass *Gerhard K.* den Kurs *Information Retrieval* hält. An diese Individualbegriffe sind dann einzelne Informationsquellen als "occurrence" angebunden. Diese Informationsquellen werden also über die Topic Map recherchierbar gemacht.

Da Topic Maps als XML-Anwendung definiert sind, ist eine Topic Map nichts anderes als ein Dokument mit einer sehr strikten und reichhaltigen Struktur. Als ASCII-Text betrachtet geht der Inhalt in der Strukturauszeichnung (XML-Tags) regelrecht unter. Abb. 4a und 4b zeigen zwei Beispiele für je ein Topic und eine Association (Beziehung), die einer öffentlich zugänglichen und zu Demonstrationszwecken vielfach verwendeten größeren XTM-Anwendung zum Thema „*Italienische Oper*“ (der Firma Ontopia AG) entnommen sind. Diese Anwendung liegt in Kapitel 5 den Abbildungen zu Grunde, die eine Topic Map als SVG-basierte interaktive Schnittstelle zeigen.

```
<association id="IDA2HPY">
  <instanceOf>
    <topicRef xlink:href="#setting"/>
  </instanceOf>
  <member>
    <roleSpec>
      <topicRef xlink:href="#opera"/>
    </roleSpec>
    <topicRef xlink:href="#carmen"/>
  </member>
  <member>
    <roleSpec>
      <topicRef xlink:href="#place"/>
    </roleSpec>
    <topicRef xlink:href="#sevilla"/>
  </member>
</association>
```

Abb. 4: Repräsentation einer exemplarischen Topic als XML-Fragment. Aus einer Topic Map zum Thema *Italienische Oper* mit ca. 640 Beziehungen (Quelle: Firma Ontopia AG)

4 Darstellung zweidimensionaler interaktiver Grafiken: SVG

Das Internet ist in den letzten Jahren tatsächlich zum multimedialen Medium geworden. Für die unterschiedlichen zu präsentierenden Datentypen gibt es eine Fülle von Formaten, wobei sich durchaus für einzelne Medientypen besonders gängige Formate herausgebildet haben: für Fotografien findet hauptsächlich das JPEG- (Joint Photographic Experts Group), für Grafiken und kleine Animationen das GIF- (Graphics Interchange Format) und für komplexe interaktive Animationen und Schnittstellen das Macromedia Flash-Format Verwendung. Dem neuen XML-

basierten Format SVG (Scalable Vector Graphics) [W3C 2001] ist zuzutrauen, in nächster Zukunft eine besondere Rolle zu übernehmen. Diese Prognose gründet sich nicht nur aus der Tatsache, dass es besonders einflussreiche Unternehmen sind, die das neue Format entwickeln und unterstützen: IBM, Microsoft, Apple, Xerox, Sun, HP, Netscape, Corel, Adobe, Quark und Macromedia. Es sind die Vorteile, die SVG gegenüber seinen Konkurrenten mitbringt:

- SVG ist ein Vektor-Format und damit skalierbar. Im Gegensatz zu den Rasterformaten jpeg oder gif bestehen Bilder also nicht aus einer Matrix einzelner Bildpunkte, sondern aus einer mathematischen Beschreibung einzelner Bildobjekte (so wird beispielsweise ein Kreis durch seinen Mittelpunkt und seinen Radius festgelegt). Dadurch kann das Rasterformat, das ein Bildschirm oder ein Drucker letztendlich für die Wiedergabe verlangt, für jede beliebige Darstellungsgröße aktuell neu und exakt berechnet werden. Die bei Rastergrafiken bekannten Treppentufen- und andere Vergrößerungseffekte treten nicht auf (siehe Abb. 5).
- SVG ist ein normales XML-Dokument. Es legt also für den menschlichen Betrachter oder etwa eine Suchmaschinen vollständig offen, aus welchen Bildobjekten es besteht, wie diese beschrieben sind und welche Texte gegebenenfalls in dem Bild vorkommen (man denke beispielsweise an eine Landkarte oder eine technische Zeichnung!). Dadurch unterscheidet es sich zum Beispiel von Macromedia Flash, das im Binärformat weder zu analysieren noch zu verändern ist. Gleichzeitig kann es wie Flash interaktiv sein. Das Hauptpotential von SVG besteht also darin, dass sich aus strukturierten Daten (z.B. relationalen oder XML-Daten) durch geeignete Transformationen bzw. Programme zweidimensionale, interaktive Schnittstellen (fast) beliebiger

Komplexität in einem offengelegten und leicht durchschaubaren Format generieren lassen.

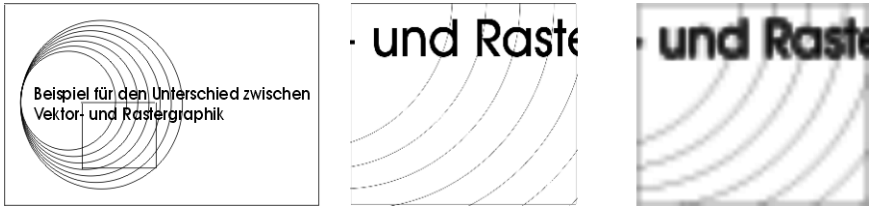


Abb. 5: Die Ausschnittsvergrößerung einer Vektorgraphik bleibt exakt (2. Bild), im Gegensatz zum Fall einer Rastergraphik (3. Bild)

So wie die Elemente einer Topic Map Topics und Associations sind, so besteht ein SVG-Dokument aus Elementen für die geometrischen Grundformen (zum Beispiel `<circle>` für den Kreis), für Texte, für Transformationen und Strukturelemente (zum Beispiel Gruppierungen von Bildobjekten). SVG stellt eine große Anzahl verschiedener grafischer Effekte (Umrandung, Füllung, Masken und Ausschnitte, Filter) zur Verfügung, insbesondere auch solche, wie sie aus der Bildbearbeitung für Rastergrafiken bekannt sind (zum Beispiel Schlagschatten). SVG bietet selbst bereits eigene Möglichkeiten der Animation von Farben, Position, der Transformation sowie weiterer Objektattribute sowie eigene Möglichkeiten der Interaktivität (Formularelemente und Ereignissteuerung) und ist darüber hinaus durch entsprechende Mechanismen für die Integration zusätzliche Elemente (foreign objects) offen.

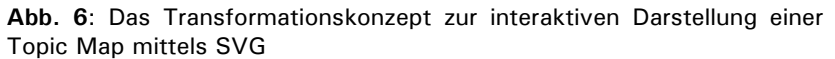
Ein konzentrierter Überblick über die Grundkonzepte und Ausgestaltung der Skalierbaren Vektorgrafiken findet sich in [Die 2001], S. 10 - 23. Im Internet wird die Bandbreite des mit SVG Möglichen an didaktisch aufbereiteten Demonstrationen zum Beispiel unter <http://www.svg-magic.com/svg-demos.html> oder <http://www.adobe.com/svg/demos/main.html> sichtbar gemacht. Besonders naheliegend kann der Vorteil von Vektor-

grafiken, den Nutzer tiefer und tiefer ins Detail hineinzoomen zu lassen, im Bereich der Kartografie ausgenutzt werden. Dies wird besonders augenscheinlich in [Win&Neu 2000] vorgeführt.

5 Von XTM zu SVG: Entwicklung eines Viewers

Die Rolle von XML als eine gemeinsame Syntax für die Integration verschiedenartiger Anwendungen wird im vorliegenden Fall besonders evident. Eine Topic Map mit einer grafisch interaktiven Schnittstelle zu versehen kann nämlich aufgefasst werden als die Transformation einer spezifischen Ausprägung von XML-Dokumenten (Dokumententyp Topic Map) in einen anderen Dokumententyp, nämlich SVG. Genau dies war die Aufgabe der Diplomarbeit [Die 2001], die in Kooperation zwischen der Fachhochschule Darmstadt, Fachbereich Informations- und Wissensmanagement, und der Abteilung e-business and information visualization des Zentrums für grafische Datenverarbeitung (ZGDV) erarbeitet wurde.

Die Architektur des entwickelten Systems ist in Abbildung 6 dokumentiert:



- der Performance
- der multiplen Wiederverwendung
- des modularen Konzeptes
- der Trennung von Inhalt und Layout

Die Transformation wird über XSLT realisiert. XSLT ist eine Teilkomponente des Standards XSL, der XML-Dokumenten ein Layout zuweisen soll. Die XSL-Transformationsregeln (XSLT) sind selbst in XML-Syntax geschrieben und haben die Aufgabe, die strukturellen Differenzen zwischen einem Originaldokument und seiner publizierten Form zu überbrücken. Mittels XSLT werden also XML-Dokumente mittels Baum-Transformationsregeln "umgebaut" und damit möglicherweise sehr

weitgehend transformiert. XSLT ist eine sehr komplexe und mächtige regelbasierte Sprache [Hol 1999]

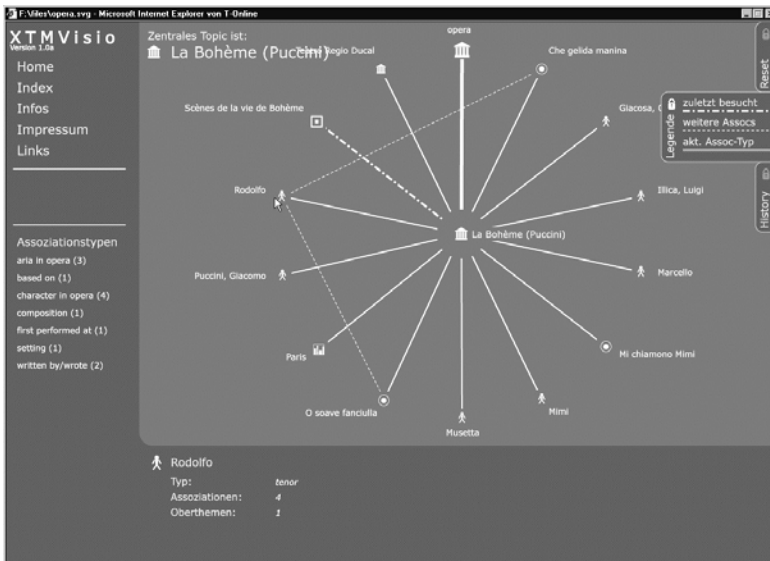


Abb. 7: Der Viewer zeigt von einem zentralen Topic aus die über Assoziationen erreichbaren Topics. Detailinformationen werden per „Mouse over“ eingeblendet (Hier zu „Tenor Rodolfo“ als textuelle Übersicht (unten) und mit seinen Querbeziehungen (als gestrichelte Linien in der Graphik)).

Ziel der ersten Umwandlungsstufe ist es, Informationen und Inhalte, die direkt oder indirekt im Quelldokument vorhanden sind, so zu organisieren, zu kontrollieren und auszubilden, dass anschließend der Zugriff und die Verarbeitung der Daten erleichtert und beschleunigt wird. Dieser Vorgang erfordert im allgemeinen viel Verarbeitungskapazität und wird a

priori berechnet. Im zweiten Umwandlungsschritt erfolgen die Erstellung und Anpassung der Anzeige- oder Verarbeitungsumgebung. Dieser Prozess kann effizient auf Anforderung angestoßen werden. Die resultierende SVG-Datei kann nun mittels eines SVG-Viewers betrachtet und erkundet werden. Das Darstellungskonzept sieht vor, dass immer genau ein Topic, das „zentrale“ Topic im Mittelpunkt steht und alle Topics (die „Sekundären“), die mit ihm über eine Assoziation in Verbindung stehen, kreisförmig angeordnet werden (Abb. 7). Um die Darstellung übersichtlich zu halten, werden Detailinformationen nur dann eingeblendet, wenn der Benutzer auf ein spezifisches Topic oder eine spezifische Assoziation zeigt. Etwas komplexer muss die Anordnung der Objekte auf dem Bildschirm dann werden, wenn Assoziationen dargestellt werden müssen, die mehr als zwei Topics miteinander verbinden (Abb. 8.).

Durch Mausklick wird ein sekundäres Topic zum neuen zentralen Topic, so dass die Topic Map schrittweise erkundet werden kann. Die im Darstellungsbereich jeweils vorkommenden Assoziationstypen werden auf der linken Seite aufgelistet und können interaktiv im Bild identifiziert werden.

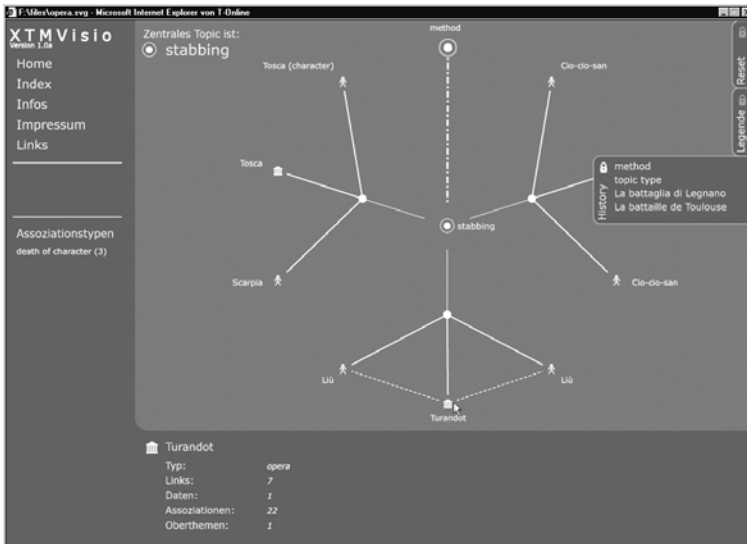


Abb. 8: Der Viewer zeigt von einem zentralen Topic aus die über Assoziationen erreichbaren Topics. Detailinformationen werden per „Mouse over“ eingeblendet.

Eine erste Inhouse-Anwendung dieses Viewers für Topic Maps wird eine Navigationshilfe in einer E-Learning-Umgebung werden. Die praktische Erprobung des Werkzeugs für einen realen Anwendungszweck steht damit unmittelbar an.

6 Ausblick

In dem vorliegenden Beitrag wurden zwei aktuelle XML-Entwicklungen zusammengebracht, die beide in verschiedener Hinsicht für die Visualisierung von Zusammenhängen eine wichtige Rolle werden spielen können: *Topic Maps* auf der Ebene der semantischen Modellierung und *Skalierbarer Vektorgrafiken* für deren Präsentation. Das Internet ist in einer solch dynamischen Entwicklung begriffen, dass man Aussagen über

dessen zukünftige Entwicklung stets mit Vorsicht begegnen sollte. Auch für die Zukunft von XTM und SVG gibt es durchaus kritische Gesichtspunkte. So scheint etwa das Verhältnis von XTM und dem Resource Description Framework (RDF) im Hinblick auf Konkurrenz oder Ergänzung noch völlig offen und SVG muss im Hinblick auf Entwicklungswerkzeuge gegenüber kommerziellen Alternativen noch sehr viel aufholen. Was die Möglichkeiten und Grenzen der in der allerletzten Zeit entwickelten und hier z.T. eingesetzten Konzepte und Werkzeuge betrifft wird erst die breite praktische Erprobung eine abgesicherte Einschätzung über deren Reichweite ermöglichen. So stellte es sich auch bei der Entwicklung des SVG-Viewers heraus, dass doch der zentrale Teil der Anwendungs- und Interaktionslogik nicht durch die XSL-Transformationen realisiert wird, sondern in den JavaScript-Anteilen der Anwendung (script.js in Abb. 6) liegt. Insofern könnte sich die Hoffnung auf „einfache“, elegante Problemlösungen auf der Basis des gegenwärtigen XML-Orchesters zum Teil wieder zerschlagen mit der Folge, dass beispielsweise das XSLT der nächsten oder übernächsten Generation nichts anderes als eine komplexe, vollständige Programmiersprache sein könnte (wie bereits gegenwärtig bei den Entwicklungsumgebungen proprietärer Produkte). Dennoch: Wenn es XTM oder SVG tatsächlich nicht sein sollten, denen die Zukunft gehört, dann werden es andere Entwicklungen sein, die viele der hier herausgestellten Eigenschaften und Vorteile ebenfalls besitzen.

Danksagung

Der Beitrag gründet weitgehend auf den zwei Diplomarbeiten [Die 2001] und [Hal 2001]. Frau Christine Haller und Herrn Alexander Diegelmann sei für ihre Arbeit großen Dank. Ebenfalls herzlichen Dank an Herrn Prof. Dr. Bernhard Thull, der mir hilfreiche Anmerkungen zu einer Vorversion dieses Artikels lieferte.

Literatur

[Bra&Lev 1985] R.J. Brachmann, H.J. Levesque: Introduction. R.J. Brachmann, H.J. Levesque: (Hr.): Readings in Knowledge Representation. Los Altos (Ca): Morgan Kaufmann 1985, S. 13-19

[Bra&Sch 1985] R.J. Brachmann, Schmolze: An Overview of the KL-ONE Knowledge Representation System, Cognitive Science 9 (1985), 2

[Car et al. 1999] S.K. Card; J.D. Mackinley; B. Shneiderman: Readings in Information Visualization - Using Vision to Think, Morgan Kaufmann Pub., 1999, S. 1-34

[COC 2000] The Apache XML Project: Cocoon – XML-based web publishing. 2000. URL: <http://xml.apache.org/cocoon/infrastructure.html>

[DäB 2000] Rolf Däßler: Trends in Visualization of Textual Information. In: G. Knorz, Ch. Otto (Hr.): Informations- und Wissensmanagement. 7. Int. Symposium für Informationswissenschaft, 8.-10.11.2000, FH Darmstadt. Tagungsband für das Anwendungsorientierte Programm. Darmstadt: FH Darmstadt, 2000, S. 105

[DäB&Pal 1998] Rolf Däßler; Hartmut Palm: Virtuelle Informationsräume mit VRML. Information recherchieren und präsentieren in 3D. Heidelberg: dpunkt, 1998

[Die 2001] Alexander Diegelmann: Visualisierung von Topic Maps mittels Scalable Vector Graphics (SVG). Diplomarbeit. FH Darmstadt, Fachbereich Informations- und Wissensmanagement, Juni 2001

[Eib 2000] Maximilian Eibl: Visualisierung im Dokument Retrieval. Theoretische und praktische Zusammenführung von Softwareergonomie und Grafik Design. Informationszentrum Sozialwissenschaften Bonn (Forschungsberichte Bd. 3), 2000

[Fav 2001] Bernhard Favre-Bulle: Information und Zusammenhang. Informationsfluß in Prozessen der Wahrnehmung, des Denkens und der Kommunikation. Wien, New York: Springer, 2001

[Gol 1990] Charles F. Goldfarb: The SGML Handbook, Oxford: Oxford University Press, 1990

[Hal 2001] Christine Haller: XML-Topic Maps für Verwaltung und Retrieval von Kompetenzen in einem F&E-Institut. Diplomarbeit. FH Darmstadt, Fachbereich Informations- u. Wissensmanagement, Juni 2001

[Hem 1993] Matthias Hemmje: eine inhaltsorientierte, intuitive 3D-Benutzerschnittstelle für Information-Retrieval-System. In: Knorz, G.; Krause, J.; Womser-Hacker, Ch. (Hr.): Information Retrieval '93. Von der Modellierung zur Anwendung. Proceedings der 1. Tagung Information Retrieval 1993. Konstanz:

Universitätsverlag Konstanz, 1993, S. 82-99

[Hol 1999] G. Ken Holman: What is xslt? 1999. URL: http://www.xslt.com/what_is.htm

[Kno 1999] Gerhard Knorz: Wissenskarten entwickeln ... Overhead-Folie zum Thema "Wissensmanagement" (Fach Informationsmethodik). WebSite ,Methodik: <http://www.iud.fh-darmstadt.de/methodik/publ/slide/wke1.htm>

[Kno 2000] Gerhard Knorz: Exemplarische Umsetzung einer Knowledge-Map. Visualisierung mittels 'klickable Imagemap' WebSite `Methodik: <http://www.iud.fh-darmstadt.de/iud/wwwmeth/publ/example/knomap1.htm>

[Koh et al 2000] T. Kohonen; S. Kaski; K. Lagus; J. Salojärvi; V. Paatero; A. Saarela:

Self Organization of a Massive Document Collection. IEEE Transactions on Neural Networks, Special Issue on Neural Networks for Data Mining and Knowledge Discovery, 11 (2000), 3, S. 574-585.

[Kol et al. 1994] W. Kollwe; M. Skorsky; F. Vogt; R. Wille: TOSCANA - ein Werkzeug zur begrifflichen Analyse und Erkundung von Daten. In: R. Wille und M. Zickwolff (Hr.), Begriffliche Wissensverarbeitung - Grundfragen und Aufgaben, B.-I. Wissenschaftsverlag, Mannheim-Leipzig-Wien-Zürich, 1994, S. 267-288.

[Min 1975] Marvin Minsky: A Framework for Representing Knowledge. In: P. Winston (Hr.): The Psychology of Computer Vision. New York: McGraw Hill, 1975

[Noh 2000] Holger Nohr: Wissen und Wissensprozesse visualisieren. Nohr, H. (Hr.): Arbeitspapiere Wissensmanagement (Stuttgart: ISSN 1616-5330), (2000), 1

[Qui 1966] Ross M. Quillian: Semantic memory. Bolt, Beranek and Newman, Cambridge, 1966

[Rat&Pep 1999] Hans Holger Rath, Steve Pepper: Topic Maps: Introduction and Allegro. 1999. URL: <http://www.topicmaps.com/content/resources/mt99/hhrstp.pdf>

[Ror&Lun 1999] Mark Rorvic; Lois F. Lunin: Introduction and Overview: Visualization, Retrieval and Knowledge. In: Journal of ASIS, 50(1999), 9, S. 790-793

[See 1997] Seemann, Patricia: A Prescription for Knowledge Management: What Hoffmann-LaRoche's Case can Teach Others. In: Perspectives on Business Innovation 1(1997),1, S. 26-33

[Tuf 1983] Edward Tufte: The Visual Display of Quantitative Information, Cheshire: Graphics Press, 1983.

[Win&Neu 2000] André M. Winter; Andreas Neumann: dynamisch generierte Kar-

togrammdiagramme. 2000. URL: www.carto.net/papers/svg/diagr.html

[W3C 1998] W3C: Extensible Markup Language (XML) Version 1.0. W3C Recommendation. In: W3C – The World Wide Web Consortium [online] 1998-02-10, english unter <http://www.w3.org/TR/1998/REC-xml-19980210.html> deutsch unter <http://www.mintert.com/xml/trans/REC-xml-19980210-de.html>

[W3C 1999] W3C: Resource Description Framework (RDF). Model and Syntax Specification

W3C Recommendation 22 February 1999. URL: <http://www.w3.org/TR/REC-rdf-syntax/>

[W3C 2001] W3C: Scalable Vector Graphics (SVG) 1.0 Specification.

W3C Recommendation 04 September 2001. URL: <http://www.w3.org/TR/SVG/>

[XTM 2000] XML Topic Maps (XTM) 1.0, 2000. URL: <http://www.topicmaps.org>

Manfred Sommer: Gemeinsame Homepages und Portale - Möglichkeiten und Voraussetzungen.

Das Internet verbindet weltweit über 100 000 000 Rechner und eine um ein Vielfaches größere Anzahl von Menschen und ist zu einer nie da gewesenen und alle Grenzen überwindenden Kommunikations-, Wissenschafts- und Wirtschaftsplattform geworden. Historisch liegen die Anfänge des Internet in dem amerikanischen ARPANET (*Advanced Research Projects Agency-Net*) das seit 1969 als militärisches Netzwerk entwickelt wurde. Ab etwa 1972 wurde es auch eingesetzt, um Universitäten und Forschungseinrichtungen zu verbinden, die mit dem Verteidigungsministerium zusammen arbeiteten. Aus Sicherheitsgründen wurde das ARPANET später in einen öffentlichen und einen nichtöffentlichen Teil getrennt. Der öffentliche Teil wurde zum Internet. In der Zeit von 1990 bis 1995 hat sich das Internet durch die Einführung des World-Wide-Web (WWW) zu einem einfach zu bedienenden Informationssystem entwickelt, und wurde damit auch für kommerzielle Anwendungen interessant. Durch die Verwendung von Grafik und verschiedenen Textformaten und der Möglichkeit, statt der strikten Menüs an beliebigen Stellen Verweise (sog. *links*) in den Text zu integrieren, ist es sofort und begeistert von der „Netzgemeinde“ aufgenommen worden und hat sich zur „Killerapplikation“ des Internets entwickelt. Firmen erkannten schnell die Möglichkeiten eines solchen universellen Kommunikationsmittels und die Öffentlichkeit ist nicht nur auf das Internet aufmerksam geworden, sondern hat das Internet mittlerweile neben Zeitung, Funk und Fernsehen als weiteres Medium akzeptiert und schätzen gelernt.

Portale

Von Anfang an existierte die Möglichkeit, Informationen in Form persönlicher Informationsseiten (*Homepages*) bereitzustellen. Recht bald kamen unpersönliche Informationsseiten zu bestimmten Themen oder für Firmen und Organisationen dazu. Heute werden diese *Portal* genannt. Einem Lexikon zufolge stammt der Begriff Portal aus dem Lateinischen und bezeichnet den monumental gestalteten Eingang eines Gebäudes. In der Antike gab es z. T. ganze Torbauten (Propyläen), auch als Ehren- und Triumphbögen. Das europäische Kirchenportal erhielt in der Zeit des romanischen Stils seine für Jahrhunderte verbindliche Form.

Im World Wide Web ist *Portal* eine Bezeichnung für viele Einstiegsseiten. Portale werden als Startseite verwendet: Benutzer können ihren Web-Browser so konfigurieren, dass er mit einem Lieblings-Portal startet. Andere Benutzer haben eine Sammlung von Lesezeichen zu verschiedenen Portalen, die sie jeweils für unterschiedliche Zwecke nutzen.

Es gibt allgemeine, sogenannte *horizontale Portale* wie z.B.

- Yahoo, Excite, Lycos, CNET, ...
- Netscape, MSN, AOL, T-Online, ...

Es gibt spezialisierte, sogenannte *vertikale Portale* wie z.B.

- Babynet.de, Haus.de, Garten.de, Garden.com, Fool.com...
- Siemens.de, Quelle.de, ...
- Geographie.de, nationalgeographic.com
- archivists.org, unesco.org/webworld/portal_archives/

Öffentliche Portale bieten Informationen und Dienstleistungen mit dem Ziel einer großen Zahl von wiederkehrenden Besuchern. Dabei werden unterschiedliche Ziele verfolgt. Ein naheliegendes Motiv für ein solches Angebot sind die zu erwartenden Einnahmen durch begleitende Darstel-

lung von Werbung. Diese Einnahmen sind umso größer, je mehr Besucher das Portal nachweisen kann. Beispiele:

- Yahoo, Excite, Lycos, CNET, ...

Firmen-Portale bieten Informationen über die jeweilige Firma an. Ein naheliegendes Motiv für ein solches Angebot ist es die jeweilige Firma bekannt zu machen und positive Informationen über die Firma zu verbreiten. Beispiele:

- Siemens, VW, ...

Eine ganze Reihe von Portalen bietet nichts anderes an als "Electronic Business". Diese *Verkaufs-Portale* stellen sich als elektronische Warenhäuser dar. Sie werden von Kaufhäusern, Banken usw. genutzt. Der Begriff *Shopping Mall* wird häufig in diesem Zusammenhang verwendet. Klassisches Beispiel ist:

- Amazon

Media-Portale bieten Zusatz-Informationen zu den jeweiligen Print-Medien. Fast alle größeren Zeitungen haben mittlerweile ein entsprechendes Angebot, auch wenn kein zusätzlicher Profit daraus resultiert. Ausschließliches Ziel: „Mitmachen ist alles“.

- Spiegel, Focus, FAZ, SZ, ...

Daneben gibt es Portale für Gruppen mit speziellen Interessen. Diese meist von privaten Personen initiierten Portale dienen der Verbreitung von Informationen zu diversen Themen.

- z.B. alle Informationen über Haus und Garten.

Es gibt *gute* und *schlechte* Portale. Besucher werten die folgenden Eigenschaften von Portalen überwiegend als positiv:

- Eine Portalseite muss schnell geladen werden.
- Wenige Bilder, kleine Bilder, viel Text.

- Hoher Nutzwert des Angebotes.
- Aktualität, Fokussierung auf eine Zielgruppe.
- Hohe Bequemlichkeit der Benutzung
- Professionelles Webdesign.

Wenn die Anbieter von Portalen die genannten Eigenschaften nicht beachten, nimmt die Nutzung meist sehr schnell ab. Wenn die Anbieter dies rechtzeitig bemerken, führt das meist sehr bald dazu, dass genau diese Regeln eingehalten werden. Anbieter müssen für einen hohen Bekanntheitsgrad und für eine positive Wertung ihres Portals sorgen. Ein weiteres wichtiges Kriterium guter Portale ist ein hoher Qualitätsstandard für die Inhalte. Nur wenn die Inhalte aktuell, zuverlässig und interessant sind, kann ein Portal auf hohe Popularität hoffen. Typische Bestandteile eines horizontalen Portals sind:

- Aktuelle allgemeine Nachrichten
- Aktuelle spezielle Nachrichten:
- Sport,
- Wetter,
- TV, Kino, etc...
- Shopping, Auktionen
- Suchmaschinen, Linksammlungen
- Chat / Foren

Ein weiteres wesentliches Angebot wird heute von den meisten Portalen realisiert: Die Personalisierung des Angebotes in Form von kostenlosen Internet- und EMail-Konten, die Speicherung persönlicher Nutzerprofile, Linksammlungen etc. Gerade der letzte Punkt ist als besonderes Angebot verschiedener Portale in letzter Zeit in Mode gekommen. Ein Lesezeichen Portal (*Bookmark Portal*) ist ein spezieller Portal- Typ, der es

dem Anwender erlaubt, seine Lieblings-Links zu speichern. Allgemein gehört dieser Typ zur Familie der *personalisierten Portal-Seiten*. Diese ermöglichen:

- Die Verwaltung von Link-Sammlungen.
- Die Verwaltung mehrerer E-Mail-Konten.
- Terminplanung und -verwaltung.
- Die Synchronisierung von PDAs.
- Das Speichern von Dateien.

Eines der ersten öffentlichen Portale wurde von Amazon.com angeboten. Dieser Klassiker startete 1995 als virtueller Buchladen. Zunächst war das ausschließliche Vertriebsziel der Verkauf von Büchern im Internet. Mittlerweile wurde der Handel ausgeweitet auf CDs, DVDs, Elektronische Geräte und vieles mehr. Filialen mit einem regional angepassten Angebot in der jeweiligen Landessprache gibt es mittlerweile in Deutschland, England, Frankreich und vielen anderen Ländern.

The screenshot shows the Amazon.de homepage. At the top, there's a navigation bar with categories like HOME, BÜCHER, MUSIK, DVD, VIDEO, ELEKTRONIK & FOTO, SOFTWARE, COMPUTER & VIDEOSPIELE, GESCHENKE & E-CARDS, AUKTIONEN, and zSHOPS. Below this is a search bar with the text "Halle, Manfred Sommer, klicken Sie hier für persönliche Empfehlungen in Bücher und Musik. Wenn Sie nicht Manfred Sommer sind oder sich abmelden möchten, klicken Sie bitte hier." The main content area is divided into several sections: "Preisbrecher: Macy Gray für 24,99 DM und noch mehr Preis-Hits Musik.", "In ALLES RUND UM...", "Für Sie Ausgewählt" (featuring "The Java Developers Almanac 2000"), "Musik" (featuring "Heather Nova komplett!"), and "Software" (featuring "Abenteuer-Trio"). On the right side, there are sections for "Aktuelles" (Business-Geschenke, Gelbe Helden!) and "Die Amazon.de Hot 100 Bücher" (Hausmannskost für Feinschmecker, Das Universum in der Fußschale).

Abb. 1: Der Klassiker "Amazon"

Als weiteres Beispiel für ein Portale wird an dieser Stelle ein virtuelles Kaufhaus vorgestellt, die Internet-Mall "Shopping24". Dieses Angebot gibt es seit Oktober 1997: Ein virtuelles Einkaufszentrum mit 21 konzerneigenen und fremden Partnern unter einer Internet-Adresse. Die Vorteile für die Mall-Anbieter sind:

- Ein gemeinsames Angebot an Service, Logistik und Werbung

Die Vorteile für die Mall-Nutzer sind:

- Ein Portal für möglichst viele Einkäufe.

shopping24 Suchbegriff eingeben **finden** shopping24-Partner **go**

[Home](#) [Detailsuche](#) [News](#) [Job-Börse](#) [Presse](#) [Hilfe](#) [Kontakt](#) [Sitemap](#)

Schnäppchen der Woche	Neu! Flugtickets	Fragen Sie Atira	Italien-Special
Nur für kurze Dauer! Attraktive Artikel zum Schnäppchenpreis.	Jetzt bei shopping24: Flüge zum Spartarif von: Frankfurt (FRA) nach: Abheben	Ihre Assistentin für alle Fälle. Fragen Sie ruhig - sie wird Ihnen gerne weiterhelfen.	Benvenuto in Italia!
Abonnieren Sie den shopping24-Newsletter und wir informieren Sie über alle Neuigkeiten.	Sie möchten auch einkaufen, wenn Sie unterwegs sind? Jetzt auch Mobil-Shopping mit dem WAP-Handy! Unser Service unter: wap.s24.de!	Machen Sie mit bei einer neuen Runde hinnwin präsentiert von sheego.com!	

Abb. 2: Die Internet-Mall "Shopping24"

Anbieter für vertikale Portale gibt es viele. Dazu zählen Berufsverbände (alles über Geographie,), Vereine (alles über Volkstanz,), Kommerzielle Nischenanbieter (Babynet,), und weitere Anbieter wie etwa die Unesco, aber auch Bibliotheken und Archive. Das gemeinsame Ziel all dieser Portale ist es, Informationen zu bestimmten Themen bzw. für bestimmte Gruppen bereitzustellen. Vertikale Portale sind themenspezifisch. Daher sind ihre Bestandteile recht unterschiedlich, auch je nach der Ebene auf der die Informationen angeboten werden:

Global:

Verweise auf untergeordnete Einrichtungen nach verschiedenen Kriterien.

Regional:

Verschiedene Gruppen von Inhalten

Mittlere Ebenen:

Verweise auf übergeordnete, untergeordnete und benachbarte Einrichtungen, aktuelle Meldungen und Nachrichten, Veranstaltungshinweise, Chat, Foren, zusammengefasste Inhalte untergeordneter Instanzen.

Lokale Ebene:

Verweise auf übergeordnete und benachbarte Einrichtungen, lokale Inhaltsangebote.

Auf allen Ebenen:

Suchmaschinen und Linksammlungen.

Austausch von Inhaltsangeboten

Bei Bibliotheken und Archiven liegt es nahe, Inhaltsangebote auf lokaler Ebene anzubieten. Ebenso naheliegend ist es, für übergeordnete Portalseiten, die Informationen der Teilnehmer zusammenzufassen. Das kann natürlich nicht *von Hand* passieren. Also müssen die Informationen in einem einheitlichen Format zur Verfügung stehen. Für diese Zwecke bietet sich aus heutiger Sicht XML an. Mit XML (**e**Xtensible **M**arkup **L**anguage) kann strukturierte Information jeder Art neutral dargestellt werden. Ziele von XML sind laut W3C (dem WWW- Consortium):

- XML wird internationales elektronisches Publizieren unabhängig vom Typ des Ziel- Mediums ermöglichen (Medien sind z.B.: gedruckt, als Folienpräsentation, als Web- Dokument, ...).
- XML wird es der Industrie ermöglichen, Plattform-unabhängige Standards für den Austausch von Daten (Stichwort EDI) zu definieren. Damit wird es zu einer Basistechnologie für Electronic Commerce.

- XML wird es ermöglichen, Informationen an Benutzer-Agenten zu übermitteln, die diese nach Erhalt automatisch verarbeiten können.
- XML kann Metadaten in Dokumenten bereitstellen - Daten über Informationen. Diese werden es ermöglichen Informationen gezielt zu suchen.

Die Struktur eines Dokumentes wird mit einer *Document Type Definition* festgelegt (DTD). Die DTD kann fehlen, im Dokument und/oder als Link angegeben sein. Eine DTD besteht aus einer Menge von „Regeln“ und beschreibt die Elemente, Attribute und Ersetzungsregeln, die in dem Dokumenttyp verwendet werden. XML erlaubt die Speicherung von Daten zusammen mit beliebigen Informationen wie Datentyp, Datendarstellung und Formatierungseigenschaften.

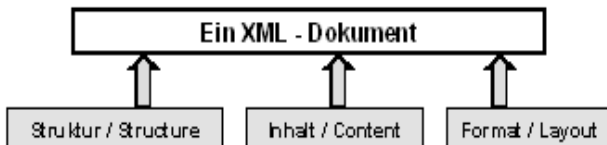


Abb. 3: Die Struktur eines XML-Dokuments

XML speichert Daten in Bäumen. Dies veranschaulicht das folgende Beispiel:

Struktur / Structure Inhalt / Content Format / Layout

Ein XML - Dokument

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Wetter>
  <Ort Messpunkt="Kirchspitze"> Harburg </Ort>
  <Datum>
    <Tag> 9 </Tag>
    <Monat> 9 </Monat>
    <Jahr> 2001 </Jahr>
    <Zeit> 09:19 </Zeit>
  </Datum>
  <Himmel> Sonnig </Himmel>
  <Temperatur Einheit="Celsius"> 42 </Temperatur>
  <Wind Einheit="km/h"> 50 10 </Wind>
  <Luftdruck Einheit="mbar"> 1033 </Luftdruck>
  <Feuchtigkeit> 90 </Feuchtigkeit>
  <Sicht Einheit="km"> 19 </Sicht>
</Wetter>
```

Abb. 4: XML: Ein Beispiel

Im Umfeld von XML ist eine ganze Reihe von zusätzlichen Standards entstanden: So ist z.B. XSLT (**e**Xtensible **S**tyle **L**anguage **T**ransformations) ein Standard zur Beschreibung von Baumtransformationen. Mit XSLT kann man verschiedene Darstellungen von Daten ineinander überführen und sie dabei z.B. für unterschiedliche Ausgabemedien anpassen. Das Ergebnis kann sein:

- ein Text mit Format-Auszeichnungen
- ein anderer XML-Baum
- ein HTML-Baum



Abb. 5: XSLT: Ein Beispiel

Allgemeiner noch sind alle möglichen Schnittstellen zu XML-Dokumenten in DOM (**D**ocument **O**bject **M**odel) beschrieben. Mit DOM-Schnittstellen kann man z.B. Java- Programme schreiben, die XML Bäume manipulieren. DOM ist eine Plattform- und Sprachneutrale Schnittstelle, die es Programmen und Skripten erlaubt, auf den Inhalt, die Struktur und den Stil eines Dokumentes zuzugreifen und ihn ggf. zu verändern. DOM enthält eine Menge von "Standard-Objekten", die HTML- und XML- Dokumente repräsentieren.

- DOM enthält ein Modell wie diese Objekte verknüpft werden können.
- DOM definiert eine Standard-Schnittstelle, um auf diese Dokumente zuzugreifen.
- Hersteller können DOM als Schnittstelle zu ihren Anwendungen anbieten und APIs um damit umzugehen. (API: Anwender Programm Schnittstelle.)

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Wetter>
<Ort Messpunkt="Kirchspitze"> Marburg </Ort>
<Datum>
<Tag> 9 </Tag>
<Monat> 9 </Monat>
<Jahr> 2001 </Jahr>
<Zeit> 09:19 </Zeit>
</Datum>
```

```
<Himmel> Sonnig </Himmel>  
<Temperatur Einheit = "Celsius"> 42 </Temperatur>  
<Wind Einheit = "km/h"> SO 10 </Wind>  
<Luftdruck Einheit = "mbar"> 1033 </Luftdruck>  
<Feuchtigkeit> 90 </Feuchtigkeit>  
<Sicht Einheit = "km"> 19 </Sicht>  
</Wetter>
```

Ausblick

XML entwickelt sich mehr und mehr zu einem neutralen Standard zur Definition und zum Austausch von Daten. Mit XML können Daten zwischen verschiedenen Anwendern, Programmen, Datenbanken, etc. ausgetauscht werden. Mit XML können Anfragen an Datenbanken, Suchmaschine und Archive gesteuert werden. XML eignet sich als Format um Daten zu beschreiben, die in HTML-Seiten eingespeist werden. XML wird zunächst HTML ergänzen und im Laufe der Zeit möglicherweise sogar ersetzen.

Index:

- access. 19, 20, 21, 22, 23, 24, 33, 42, 198
AIDA..... 129, 133
Anwendertests..... 197
Archivalientypen..... 154
Archivschule 7, 8, 9, 14, 53, 83, 115, 116, 135, 153, 180
Austauschformat 118, 119, 126, 127, 147, 154, 198, 204
Banken 183
Behörden 97, 103, 139
Benutzer 9, 11, 13, 14, 16, 84, 102, 115, 116, 118, 123, 129, 137, 139, 144, 145, 146, 156, 164, 166, 175, 177, 190, 192, 193, 194, 213, 220, 227
Benutzeroberfläche 104, 106
Benutzerorientierung 94
Benutzerschnittstellen 196
Beratung..... 129, 240
Beständebaum..... 94
Beständeübersicht .. 11, 12, 14, 15, 67, 86, 95, 115, 116, 159, 176
Bewertung 98, 100, 102, 103, 106, 176
Bildbearbeitung..... 71, 81, 209
Binary large objects..... 148
CLIR..... 22, 52
collection..... 28, 29, 31, 34, 58
csv 162
Datenbank . 56, 57, 61, 63, 71, 84, 85, 87, 88, 95, 99, 108, 109, 110, 112, 117, 138, 143, 148, 154, 155, 156, 160, 162, 163, 164, 165, 170, 183, 185, 186, 188
Datenvolumen 146
DFG 51, 52, 68, 83, 112, 115, 118, 173, 174, 177, 178, 180, 240
Dienstleister 76, 80, 83, 84
Digitalisierung... 67, 68, 69, 70, 76, 77, 78, 80, 84, 118, 120, 150, 179
DOM 229
DTD 37, 39, 44, 45, 126, 149, 227
Dublin Core 126
EAD 7, 21, 23, 27, 29, 31, 33, 34, 37, 39, 40, 41, 42, 43, 44, 45, 46, 49, 50, 51, 52, 53, 54, 55, 56, 57, 59, 60, 61, 63, 64, 65, 111, 112, 126, 127, 147, 171, 180
Entstehungszusammenhang 109
Erschließung..... 14, 70, 73, 74, 98, 102, 104, 105, 106, 110, 112, 113, 125, 135, 136, 137, 139, 142, 143, 148, 149, 176
Fachanwendung 139
Fachkonzept..... 49, 51, 135, 147, 185, 190
Faksimiles 88
Farbmanagement 78
Farbmikrofilm 76
Filesystem..... 84, 85, 86
Findbuch.... 13, 14, 15, 70, 83, 85, 86, 87, 88, 89, 90, 91, 98, 102, 108, 123, 129, 159, 171, 193
Findbuchdaten..... 118, 126
Findbücher... 7, 11, 12, 14, 15, 16, 50, 67, 68, 72, 95, 98, 99, 100, 108, 109, 111, 112, 115, 116, 117, 118, 119, 120, 121, 122,

124, 125, 126, 127, 142, 148,
154, 163, 167, 170, 171, 177
Findmittel.. 12, 49, 50, 54, 56, 57,
63, 64, 70, 87, 111, 115, 117,
121, 125, 129, 145, 150, 155,
162, 190, 191
Frames.....203
Funktionen 58, 123, 142, 206
Generator85
Generierung..... 85, 116, 117
Geschichte 97, 98, 177
GIF.....79, 207
Gliederung..... 11, 15, 158, 161
Gliederungspunkt 164, 170
Globalisierung.....181
Graustufen 77, 78, 79
HADIS ...116, 126, 153, 154, 157,
160, 161, 164, 169
HTML29, 34, 85, 86, 147, 199,
228, 229, 230
Identifikation80
Informationsdienstleister173
Interaktion..... 10, 178
ISAD(G) 30, 34, 37
JPEG 79, 207
Klassifikation ... 87, 100, 102, 106,
120, 122, 125, 142, 163, 194
Kompatibilität 104, 126
Komprimierung79
Konvertierung 79, 122
Kooperation... 119, 135, 151, 174,
178, 179, 210
Laufzeit....58, 123, 158, 163, 169,
194
Layout 123, 197, 211, 227
Lesezeichen.....220, 222
LINUX.....94
Magazinverwaltung 102, 137, 139,
140

MALVINE 23
MARC..... 29, 30, 32, 34
MARC-AMC..... 20
Massenanwendung..... 184
Medienbruch..... 174
Merkzettel 170
Metadaten 70, 71, 74, 80, 227
Metadatenkatalog..... 132, 134
MIDOSA.... 73, 75, 83, 85, 86, 87,
88, 89, 90, 91, 95, 108, 116,
126
Mikrofilm.....69, 70
Nachrichten 183
Navigation . 85, 86, 168, 187, 188,
193, 201, 204
Nutzer67, 89, 175, 177, 178, 180,
195, 200, 210, 224
Nutzergruppe 11
Nutzung 7, 9, 11, 13, 15, 16, 123,
135, 140, 154, 162, 179, 181,
195, 197, 201, 222
Objekt..... 70, 148, 157, 158, 159,
161, 166
offline 67
Orientierung..... 13, 123, 145, 195
PARSIFAL.....83, 95
Portal..... 183
Präsentation 7, 50, 55, 56, 68, 70,
71, 75, 76, 82, 83, 87, 93, 94,
95, 98, 108, 111, 112, 115,
116, 130, 176, 214
Präsentationsmodelle 69
Provenienz..... 96
Qualitätskontrolle78, 80, 81, 82
Rasterdaten 132
Reproaufträge 185
Retrodigitalisierung 50

- Retrokonversion 112, 115, 117,
119, 120, 121, 122, 124, 125,
127, 147
- RLG 7, 19, 20, 21, 23, 25, 37, 44,
56
- Sachaktenverzeichnung 143
- Scannen 76, 77, 125
- Schnittstelle... 108, 126, 139, 147,
162, 179, 198, 206, 207, 208,
210, 229
- Serien 106, 142, 163
- series .. 28, 39, 40, 41, 42, 46, 58,
60
- Server ... 71, 84, 85, 94, 143, 146,
155, 164, 167, 175, 184, 186,
211
- SGML..... 54
- Standards .. 19, 28, 30, 32, 50, 53,
104, 120, 155, 179, 211, 226,
228
- Standardsoftware 71, 83, 178
- Strukturansicht.. 68, 85, 86, 87, 88
- Strukturbaum 130, 131, 134
- Strukturen . 7, 9, 11, 16, 111, 113,
122, 142, 174, 178
- Suchergebnisse 87, 195
- Suchmaschine 7, 9, 86, 230
- SVG 198, 206, 207, 208, 209,
210, 211, 213, 215, 216, 218
- Syntax 204, 210, 211, 218
- Tabellen 101, 102, 123, 148
- Tags 60, 64, 127, 206
- Tektonik 11, 15, 68, 130
- Textdatei 117
- Themenkarten 199, 200
- training 21, 25, 39, 40, 41, 46
- Transkription 68, 88
- Übergabe 103, 154
- Übersichtlichkeit 203
- UNIX 34, 143
- V.E.R.A. 116, 135, 140, 141, 142,
145, 146, 147, 148, 149, 150
- Verwaltungsakten 58
- Visualisierung ... 72, 154, 195, 196,
198, 200, 201, 214, 216, 217
- Vorschaubild 89, 90
- Wavelet 79
- Werkzeuge. 67, 68, 78, 83, 87, 95,
117, 121, 126, 149, 150, 180,
197, 215
- Wissenskarten 198, 199, 217
- Wissensmanagement..... 195, 197,
198, 210, 216, 217
- XML 7, 38, 39, 44, 46, 54, 55,
108, 127, 146, 147, 148, 149,
155, 156, 160, 170, 171, 197,
198, 204, 206, 207, 208, 210,
211, 214, 216, 218, 226, 227,
228, 229, 230
- XSLT 211, 215, 228, 229
- Ziel . 12, 13, 76, 77, 99, 102, 106,
112, 118, 124, 126, 127, 156,
166, 174, 177, 179, 212, 220,
221, 225, 226
- Zugang 53, 59, 67, 94, 99, 102,
112, 125, 129, 130, 134, 145,
155, 156, 160, 161, 164, 166,
192, 240
- Zugänglichkeit 68, 115
- Zukunft 7, 113, 162, 198, 208,
215

Die Autorinnen und Autoren

Paul Bantzer
Geschäftsführer
startext Unternehmensberatung GmbH
Kennedyallee 2
53175 Bonn
Telefon 02 28/ 95 99 6-0
Email: pba@startext.de

Dr. Frank M. Bischoff
Wissenschaftlicher Angestellter
Staatsarchiv Münster
Bohlweg 2
48147 Münster
Telefon: 0251/4885-0
Email: frank.m.bischoff@stams.nrw.de

Dr. Mechthild Black-Veldtrup
Wissenschaftliche Angestellte
Hauptstaatsarchiv Düsseldorf
Mauerstr. 55
40476 Düsseldorf
Telefon: 0211-9449-02
Email: mechthild.black-veldtrup@hsa.nrw.de

Dr. Manfred von Boetticher
Archivdirektor
Niedersächsisches Hauptstaatsarchiv
Am Archiv 1
30169 Hannover
Telefon: 0511-120-6601
Fax: 0511-120-6699
Email: Manfred.Boetticher@staatsarchiv-h.niedersachsen.de

Dr. Edgar Büttner
Archivoberrat
Bundesarchiv
Potsdamer Str. 1
56064 Koblenz
Telefon: 0261-505-260
Email: e.buettner@barch.bund.de

Anne van Camp
Manager, Member Initiatives
Research Libraries Group
1200 Villa Street
Mountain View, CA 94041-1100 USA
Telefon: 650-691-2237
Email: Anne.Van.Camp@notes.rlg.org

Catherine Dhérent
conservateur général du patrimoine, chargée de mission auprès de la
Directrice des Archives de France pour les technologies de l'information
et de la communication
56 rue des Francs-Bourgeois
75141 Paris Cedex 03
Telefon : 01 40 27 62 65
Email: catherine.dherent@culture.gouv.fr

Dr. Sigrun Eckelmann
Programmdirektorin
Wissenschaftliche Literaturversorgungs- und Informationssysteme 4
Deutsche Forschungsgemeinschaft
Kennedyallee 40
53175 Bonn
Telefon: 0228/885-2344
Email: sigrun.eckelmann@dfg.de

Thomas Fricke
Wissenschaftlicher Angestellter
Landesarchivdirektion Baden-Württemberg
Eugenstr. 7
70182 Stuttgart
Telefon: 0711-2124272
Fax 0711-2124283
Email: Maier@lad-bw.de

Dr. Bernhard Grau
Archivrat
Bayerisches Hauptstaatsarchiv
Schönfeldstraße 5-11
80539 München
Telefon: 089/28638-2575
Fax 089/28638-2954
Email: bernhard.grau@bayhsta.bayern.de

Bernward Helfer
Oberamtsrat
Hessisches Hauptstaatsarchiv
Mosbacher Str. 55,
65187 Wiesbaden
Telefon: 0611/881-0
Email: B.Helfer@hhstawi.hessen.de

Kris Kiesling
Head, Department of Manuscripts & Archives
Harry Ransom Humanities Research Center
P.O. Drawer 7219
University of Texas at Austin
Austin, TX 78713-7219
Telefon: (512) 232-4614
Fax: 512.471.2899
Email: kiesling@mail.utexas.edu

Prof. Dr. Gerhard Knorz
Fachhochschule Darmstadt
Fachbereich Informations- und Wissensmanagement
Schöffersstr. 3, Hochhaus Raum 714 (Postadresse: Haardtring 100)
64295 Darmstadt
Telefon: 06151-16-8499
Fax: 0651-16-8980
E-Mail: knorz@iud.fh-darmstadt.de

Dr. Gerald Maier
Archivrat
Landesarchivdirektion Baden-Württemberg
Eugenstr. 7
70182 Stuttgart
Telefon: 0711-2124272
Fax: 0711-2124283
Email: Maier@lad-bw.de

PD Dr. Angelika Menne-Haritz
Ltd. Archivdirektorin
Archivschule Marburg
Bismarckstr. 32
35037 Marburg
Telefon: 06421-1697113
Email: mennehar@mail.uni-marburg.de

Prof. Dr. Manfred Sommer
Philipps Universität Marburg
Fachbereich Mathematik und Informatik
Fachgebiet Informatik
Hans-Meerwein Str.
35032 Marburg
Telefon: 06421-2825430
Email: sommer@mathematik.uni-marburg.de