



# .hist 2006

## Geschichte im Netz: Praxis, Chancen, Visionen

ISSN: 1612-5940

# Historisches Forum

10 • 2007 • Teilband I

Vom 22. bis zum 24. Februar 2006 veranstaltete der Kooperationsverbund Clio-online in Berlin die Tagung „.hist 2006: Geschichte im Netz – Praxis, Chancen, Visionen“. Dies war bereits der zweite große Kongress an der Humboldt-Universität zu Berlin und der Berlin-Brandenburgischen Akademie der Wissenschaften, der sich mit „Geschichte im Netz“ und mit den Möglichkeiten und Grenzen der Nutzung von Kommunikationsnetzwerken durch die Geschichtswissenschaft in theoretischer Reflexion wie praxisbezogener Diskussion beschäftigte. Der Austausch zwischen Geisteswissenschaftlern/innen, die mit netzbasierten Medien arbeiten, Archivaren/innen, Bibliothekaren/innen, Museumsmitarbeitern/innen und Informatikern/innen zielte auf die Überwindung disziplinärer und institutioneller Grenzen, was sich auch in den vorliegenden Bänden widerspiegelt.

Die Konferenz bot in zwölf Sektionen und Workshops, zwei Podiumsdiskussionen und an zahlreichen Projektständen mit insgesamt etwa hundert Vorträgen, Diskussionsrunden und Präsentationen mehr Informationen, als die zahlreichen Besucher/innen aufnehmen konnten. Umso mehr freuen wir uns, im vorliegenden Doppelband schriftliche Ausarbeitungen der meisten Referate, bereits über H-Soz-u-Kult veröffentlichte Querschnittsberichte zur Tagung sowie eine Zusammenfassung der Abschlussdiskussion vorlegen zu können. Die Teilbände 1 und 2 des Historischen Forums Nr. 10 bündeln mit insgesamt über 80 Beiträgen die Mehrzahl der auf der Tagung „.hist 2006“ präsentierten Vorträge und machen sie allen Interessierten zugänglich.

ISBN: 978-3-86004-205-2

## Beiträge der Tagung .hist 2006

Herausgegeben für Clio-online von  
Daniel Burckhardt, Rüdiger Hohls und Claudia Prinz



[http://edoc.hu-berlin.de/e\\_histfor/10\\_I/](http://edoc.hu-berlin.de/e_histfor/10_I/)  
Veröffentlichungen von Clio-online, Nr. 2



DIGITALISIERUNG UND STRATEGIEN  
DER LANGZEITARCHIVIERUNG

DIE DEUTSCHE FORSCHUNGSGEMEINSCHAFT UND  
DEUTSCHE INITIATIVEN ZUR LANGZEITARCHIVIERUNG  
FÜR DIE WISSENSCHAFT

von Max Vögler

*Die Erschließung und Bereitstellung historischer Dokumente ist ein etablierter Förderschwerpunkt der Deutschen Forschungsgemeinschaft. Aber es ist ein Förderschwerpunkt im Umbruch: die neuen Möglichkeiten im Bereich des digitalen Publizierens haben die Arbeitsweise von Historikern/innen radikal verändert und stellen auch neue Anforderungen an geschichtswissenschaftliche Infrastrukturen. Dieser Beitrag befasst sich vor allem mit den strukturellen Aspekten dieser Entwicklung auf nationaler Ebene und fragt hier insbesondere nach den Projekten, Perspektiven und Problemen im Bereich der Langzeitarchivierung digitaler Dokumente.*

\*\*\*

Als im Jahre 1086 Wilhelm der Eroberer das *Domesday-Projekt* ins Leben rief, wollte er eine für alle Zeit nachvollziehbare und *endgültige* Darstellung der damaligen Grundbesitz-Verhältnisse in England schaffen. Dies ist ihm gelungen: Fast ein Jahrtausend später gilt das so genannte *Domesday-Buch* immer noch als erster umfassender Überblick über Grundbesitz in England.<sup>1</sup> 1986 hat die BBC zum 900. Jubiläum des Buches einen ähnlichen Ansatz verfolgt. Das *BBC Domesday-Projekt* sollte einen Querschnitt der „images and sounds of Britain 1986“ – insgesamt 200.000 digitale Dokumente (Ton, Film, Text, Bild, usw.) – als kulturelle Überlieferung für das nächste Jahrtausend sichern. 2,5 Millionen britische Pfund hat das Projekt gekostet und die Ergebnisse wurden auf besonders unempfindliche

---

1 Vgl. <<http://www.domesdaybook.co.uk/>> (11.08.2006).

Laserdisketten gebrannt. Allerdings konnten die Laserdisketten nur auf einem Acorn special BBC Microcomputer gelesen werden und würden demnach nur so lange verfügbar sein, wie das letzte Exemplar des Acorns noch funktionsfähig blieb. Schon wenige Jahre später war die Technik obsolet, die Inhalte unlesbar. Erst 2002, nach dreijähriger Arbeit, ist es Wissenschaftlern/innen an der University of Michigan im so genannten *Camelion-Projekt* gelungen, eine Emulationsumgebung für das Acorn Betriebssystem zu programmieren, die die Inhalte des *BBC Domesday-Projektes* wieder zugänglich macht. Der Projektmanager bemerkte 2002 zu Recht, „*BBC Domesday* has become a classic example of the dangers facing our digital heritage.“<sup>2</sup>

Die zunehmende Verwendung von digitalen Inhalten – ob *born digital*, wie zum Beispiel E-Mails oder Webseiten, oder nachträglich digitalisiert – im wissenschaftlichen Alltag ist für Bibliotheken, Archive, Rechenzentren, Behörden, Verlage und auch für die Wissenschaftler/innen selbst eine enorme Herausforderung. Die ständig sinkenden Kosten der Digitalisierung und Optical Character Recognition-Kodierung, die neuen Möglichkeiten der semantischen Erschließung, Tagging und anderen Entwicklungen sind dabei, auch die traditionelle Arbeitsweise der Historiker/innen – in Archiven und Bibliotheken handgeschriebene oder gedruckte Quellen zu lesen und darüber Aufsätze und Monografien zu schreiben – radikal zu verändern. In den Sprach- und Literaturwissenschaften sowie in der Philosophie gibt es schon die ersten Ansätze einer neuen, kollaborativen Arbeitsweise.<sup>3</sup>

Dieser Beitrag bietet eine kurze Einführung in die Probleme und Perspektiven der Langzeitarchivierung. Es werden anfänglich allgemeine Betrachtungen zur Langzeitarchivierung gemacht, anschließend einige Entwicklungen in Deutschland vorgestellt und letztlich die momentanen Überlegungen, Fragestellungen und Förderperspektiven der DFG dargestellt.

2 Zum Domesday Projekt vgl. <<http://news.bbc.co.uk/1/hi/technology/2534391.stm>>; <<http://www.si.umich.edu/CAMILEON/>> (11.08.2006).

3 Hierzu vgl. z.B. das Projekt Hyper-Nietzsche: <<http://www.hypernetzsche.org/>> (11.08.2006).

### Warum Langzeitarchivierung?

Das Ziel der Langzeitarchivierung ist schnell formuliert, aber tückisch in der Umsetzung. Im Gegensatz zur „herkömmlichen“ Langzeitarchivierung – die Aufbewahrung eines Buches oder Schriftstücks in einer Bibliothek oder einem Archiv – muss bei der Langzeitarchivierung digitaler Objekte nicht nur auf den Erhalt des Inhaltes geachtet werden, also auf die Integrität der digitalen Information, sondern auch darauf, dass diese Inhalte verwendbar bleiben. Digitale Medien langfristig zu archivieren, bedeutet also nicht in erster Linie das Speichern, sondern die dauerhafte Nutzung zu ermöglichen, so dass Texte, Bilder, E-Learning-Objekte (Leselisten, virtuelle Semesteraperrate, Projektergebnisse der Studierenden), digitale Reproduktionen von Museumsstücken, Datenbanken, sowie Forschungsdaten und jegliche andere Art digitaler Information über viele Jahre und Softwareversionen hinweg verwendbar bleiben.

Grundsätzlich gibt es zwei fast entgegengesetzte Strategien, die jeweils die dauerhafte Nutzung digitaler Objekte zum Ziel haben. Die erste ist eine Migrationstrategie, bei der die regelmäßige Überspielung von digitalen Inhalten auf das jeweils gängige Format im Mittelpunkt steht. Die zweite ist eine Emulationsstrategie, bei der das Daten- oder Softwareformat erhalten bleibt und die Nutzung durch eine in regelmäßigen Abständen angepasste Emulationsumgebung dauerhaft ermöglicht wird. Dahinter stehen zwei konträre Philosophien: Was soll erhalten bleiben, Inhalt oder Objekt? Marilyn Geller hat dieses Dilemma mit all seinen Konsequenzen gut beschrieben:

„If it is the object we want to preserve, then we must choose emulation, which requires the building of backward compatible software. If instead, we want only to preserve the intellectual content, then the more important issue is to normalize files in some way that will allow us to store and manipulate them to assure readability in a future software and hardware environment.“<sup>4</sup>

Zu diesen technischen Herausforderungen kommen weitere. Die naht-

4 Vgl. Geller, Marilyn, Models for E-Journal Archives. Future Pathways into the Past, The E-Resources Management Handbook, Bd. 1 (2006), S. 52-53.

lose Einbindung *rechtlicher* Regelungen bei automatisierten Zugriffssystemen ist noch weitgehend unerprobt. Hierzu müssen Verlage aktiv beteiligt werden, wenn es um die Speicherung von lizenzierten Inhalten geht. Archive, in denen der Zugang zu (digitalisierten) Beständen zunehmend über das Internet gewährleistet wird, werden robuste und automatisierte Datenschutzregelungen implementieren müssen. Weiterhin kann eine Langzeitarchivierungsstrategie nur im Netzwerk und in Zusammenarbeit mit vielen Partnereinrichtungen vollzogen werden. Das heißt, dass *organisatorische* Modelle – Workflows der Datenab- und -weitergabe, Bearbeitung, Aufbau von Speicherungsnetzwerken durch Datenspiegelung usw. – auch wichtige Bestandteile einer Archivierungsstrategie sind. Und verbunden mit dieser organisatorischen Herausforderung gibt es die *finanzielle* Komponente: Wie können Geschäftsmodelle erarbeitet werden, die die Kosten der Langzeitarchivierung in einer handhabbaren Weise im System verteilen?

Diese vier Problemkomplexe der Langzeitarchivierung – Technik, Organisation, Recht und Finanzierung – müssen auf lokaler, regionaler sowie nationaler und internationaler Ebene angegangen werden. Sie müssen auch für verschiedene Arten von digitalen Objekten – E-Journals, E-Books, Forschungsprimärdaten aus verschiedensten Disziplinen, Digitalisate, usw. – umgesetzt werden. Nur so kann ein wirklich „vertrauenswürdiges Netzwerk“ für die Langzeitarchivierung digitaler Objekte geschaffen werden.

### Lösungsansätze in Deutschland

In Deutschland bilden zwei Projekte den Kern einer bundesweiten Langzeitarchivierungsstrategie. Vor allem aufbauend auf die Vorarbeiten der Kollegen/innen an den Nationalbibliotheken in den Niederlanden und in Australien<sup>5</sup> verfolgt das deutsche Projekt *kopal* die Entwicklung einer auf die hiesigen Bedürfnisse angepassten technischen Lösung für die Langzeitarchivierung.<sup>6</sup> Gleichzeitig zielt das Projekt *nestor – Kompetenznetzwerk*

5 Vgl. <<http://www.kb.nl/dnp/e-depot/e-depot-en.html>> sowie <<http://www.nla.gov.au/preserve/>> (11.08.2006).

6 Das Projekt ist an der Deutschen Nationalbibliothek in Frankfurt sowie der Gesellschaft für wissenschaftliche Datenverarbeitung mbH und der Staats- und

*Langzeitarchivierung* darauf ab, Konsens über eine dauerhafte Organisationsform sowie die Abstimmung über die Übernahme von Daueraufgaben herzustellen.<sup>7</sup>

Weitere Maßnahmen sind jedoch auf verschiedenen Ebenen und mit unterschiedlichen Ansätzen notwendig, um ein wirklich funktionierendes System der digitalen Langzeitarchivierung zu schaffen. Die Deutsche Forschungsgemeinschaft (DFG) hat als zentrale Selbstverwaltungseinrichtung der Wissenschaft ein besonderes Interesse an dem Erhalt, der Pflege und der Gewährleistung des langfristigen Zugriffs auf wissenschaftlich relevante digitale Inhalte. Die Förderung von Langzeitarchivierungsstrategien ist der DFG bei den nachfolgend aufgeführten drei Segmenten wissenschaftlicher Informations-Infrastrukturen besonders wichtig:

1. Das System der überregionalen Literaturversorgung: Sondersammelgebietsbibliotheken (SSG-Bibliotheken) sammeln zunehmend elektronische Dokumente in Form von E-Books, E-Journals, Webseiten, pdf-Dokumenten usw. Für diese Materialien sollen sie „die über die Beschaffung hinausgehenden Funktionen der professionellen bibliothekarischen Pflege der Sammlungen von der Erschließung über die überörtliche Bereitstellung der Dokumente bis zur Sicherung der langfristigen Verfügbarkeit“ gewährleisten.<sup>8</sup>
2. Digitalisierungsprojekte: Im Rahmen von (Retro-)Digitalisierungsprojekten wurden seit 1996 mit erheblichem Einsatz von Fördermitteln „Primärdaten der geisteswissenschaftlichen Forschung“ online verfügbar gemacht.<sup>9</sup>

---

Universitätsbibliothek in Göttingen (SUB) angesiedelt und wird durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert. Vgl. <<http://kopal.langzeitarchivierung.de/>> (11.08.2006).

7 Nestor ist auch in Göttingen an der SUB angesiedelt und wird durch das BMBF finanziert. Vgl. <<http://www.langzeitarchivierung.de/>> (11.08.2006).

8 „Richtlinien zur überregionaler Literaturversorgung der Sondersammelgebiete und virtuellen Fachbibliotheken“ (Stand 17.07.2006), S. 6. Vgl. <[http://www.dfg.de/forschungsfoerderung/wissenschaftliche\\_infrastruktur/lis/download/richtlinien\\_lit\\_versorgung\\_ssg\\_0607.pdf](http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/richtlinien_lit_versorgung_ssg_0607.pdf)>. Zum System der SSG-Bibliotheken vgl. <<http://webis.sub.uni-hamburg.de/>> (11.08.2006).

9 Hierzu gehören die DFG-geförderten Digitalisierungsprojekte (für eine Übersicht vgl. <<http://www.zvdd.de/sammlungen.html>>); die zunehmende Zahl der archi-

3. Forschungsprimärdaten: DFG-geförderte Forschungsprojekte in allen Disziplinen produzieren große Mengen an Primärdaten. Hier sind Beiträge zu einer kohärenten Strategie zu leisten, um die Speicherung, das Auffinden und den dauerhaften Zugang von Forschungsdaten zu ermöglichen.

Im Rahmen dieser Handlungsfelder hat die DFG in den vergangenen Jahren eine Reihe von Projekten gefördert, die weniger den Gesamtanspruch von *nestor* oder *kopal* verfolgen, sondern vielmehr einen Beitrag zu Teilaspekten und -fragen leisten. Hier sind als Projekte hervorzuheben:

- *Datenbankgestützte Langzeitarchivierung digitaler Objekte*, Universität der Bundeswehr München: allgemeine Konzeptentwicklung für die Überführung digitaler Publikationen in Standarddatenbanksysteme.<sup>10</sup>
- *BABS: Langzeitarchivierung von Netzpublikationen aus dem Spektrum der Bayerischen Staatsbibliothek (BSB)*, BSB München: organisatorische und rechtliche Musterlösungen für elektronische Medien aus Verlagen und Behörden.<sup>11</sup>
- Erfassung, Erschließung und Sicherung von Websites politischer Parteien der Bundesrepublik Deutschland sowie ihrer Fraktionen in den Parlamenten, Archiv der Friedrich-Ebert-Stiftung Bonn: Entwick-

---

varischen Findmittel (und z.T. auch Digitalisate), die online auffindbar sind (vgl. <[http://www.bundesarchiv.de/bestaende\\_findmittel/findmittel\\_online/index.html](http://www.bundesarchiv.de/bestaende_findmittel/findmittel_online/index.html)>; <<http://www.archive.nrw.de/>>; <<http://www.landesarchiv-bw.de/>> sowie <<http://kalliope.staatsbibliothek-berlin.de/>> für Nachlässe und Autografen); die DFG-geförderten Nationallizenzen für die Wissenschaft, in denen auch viele für die Geisteswissenschaften relevanten Quelleneditionen enthalten waren (z.B. „Early English Books“, Comintern Classified Archives, oder das Patrologia graeco-latina; hierzu vgl. <<http://www.nationallizenzen.de/>>); die weiteren Digitalisierungsprojekte in der EU und der restlichen Welt (hier z.B. die vielen Digitalisierungsprojekte des Joint Information Systems Committee (JISC), aufgefasst unter: <[http://www.jisc.ac.uk/index.cfm?name=coll\\_subject\\_a](http://www.jisc.ac.uk/index.cfm?name=coll_subject_a)> sowie das gigantische „American Memory“ Projekt der Library of Congress (LOC), vgl. <<http://memory.loc.gov/ammem/index.html>>); und noch die nicht-staatlichen not-for-profit Digitalisierungsprojekte, vor allem im Bereich der Zeitschriften (vgl. <<http://www.jstor.org/>> und <<http://www.digizeitschriften.de/>>) (11.08.2006).

10 Vgl. Borghoff, Uwe M.; Rödiger, Peter; Scheffczyk, Jan; Schmitz, Lothar, Langzeitarchivierung. Methoden zur Erhaltung digitaler Dokumente, Heidelberg 2003.

11 Vgl. <<http://www.babs-muenchen.de/>> (11.08.2006).

lung und Erprobung von Verfahren zur automatisierten Spiegelung von Websites.<sup>12</sup>

- *Wissenschaftliche Primärdaten in der Meteorologie*, Technische Informationsbibliothek (TIB) Hannover: Bibliothek als Vermittlungsdienst – eine DOI (Digital Object Identifier) Vermittlungsstelle – zu wissenschaftlichen Primärdaten.<sup>13</sup>
- *Dokumentation und Archivierung von Rohdatensätzen aus der psychologischen Forschung*, Zentrum für Psychologische Information und Dokumentation (ZPID) Trier: gesicherte Archivierung durch Metadaten genau beschriebener Rohdatensätze zwecks Re- und Sekundäranalysen durch Dritte.<sup>14</sup>
- *Pressearchiv des Hamburger Weltwirtschaftsarchivs (HWWA)*, HWWA Hamburg: sukzessive Freischaltung von urheberrechtlich geschützten Digitalisaten.<sup>15</sup>

Als nächsten Schritt plant die DFG die gezielte Förderung von verteilten Speichersystemen für digitale Inhalte. So können die Integrität der je gespeicherten Daten systemseitig permanent überprüft und gegebenenfalls auftretende Verluste selbsttätig kompensiert werden. Eine solche verteilte Speicherung bietet sich zum Beispiel für die Inhalte von Open Access-Zeitschriften, Materialien aus den Projekten zur retrospektiven Digitalisierung von Bibliotheksbeständen sowie die Inhalte der Fachinformationsführer an.

Darüber hinaus gibt es aber noch weitere Ansätze und Fragestellungen, die wiederholt in verschiedenen Kontexten der wissenschaftlichen Datenhaltung erprobt bzw. beantwortet werden müssen:

1. *Technik*: Welche bereits verfügbaren technischen Lösungen können für die Langfristarchivierung im jeweiligen Handlungsfeld (SSG-Bibliotheken, Digitalisierungsprojekte, Forschungsprimärdaten) eingesetzt werden? Welche Vor- und Nachteile resultieren aus dem Einsatz verschiedener Lösungen?

---

12 Vgl. <<http://www.fes.de/archiv/spiegelung/default.htm>> (11.08.2006).

13 Vgl. <[http://www.std-doi.de/front\\_content.php](http://www.std-doi.de/front_content.php)> (11.08.2006).

14 Vgl. <<http://www.zpid.de/>> (11.08.2006).

15 Vgl. <<http://webopac.hwwa.de/digiview/>> (11.08.2006).

2. *Organisation*: Wie kann es gelingen, Netzwerke von vertrauenswürdigen Partneereinrichtungen aufzubauen? Unter welchen Voraussetzungen und aufgrund welcher organisatorischen Vorbedingungen können die im System der überregionalen Literaturversorgung engagierten Bibliotheken ein derartiges Netzwerk bilden? Nach welchen Modellen wären die langfristige Verfügbarkeit und der dauerhafte Zugang zu den Ergebnissen von Digitalisierungsprojekten verlässlich zu garantieren? Mit welchen Modellen könnte das Auffinden von und der Zugriff auf wissenschaftliche(n) Forschungsprimärdaten langfristig und gegebenenfalls über die Grenzen einzelner Disziplinen hinweg ermöglicht werden?
3. *Recht*: Welche rechtlichen Regelungen sind bei der Langzeitarchivierung lizenzpflichtiger und urheberrechtlich geschützter Inhalte zu beachten? Inwieweit sind auch datenschutzrechtliche Aspekte zu bedenken? Wie können etwa maschinell lesbare Rechtssysteme, oder die Zusammenarbeit mit Verlagen in Public-Private-Partnership-Modellen in diesem Bereich zu Problemlösungen beitragen? Und, last but certainly not least,
4. *Kosten*: Mit welchen Kosten ist beim Aufbau eines Langzeitarchivierungs-Netzwerks in verschiedenen Handlungsfeldern zu rechnen? Wie skalieren sich diese Kosten beim nachhaltigen und dauerhaften Betrieb dieser Systeme mit mehreren Partnern? Welche Geschäftsmodelle können dazu beitragen, vertrauenswürdige und tragfähige Netzwerke für die Archivierung digitaler Objekte aufzubauen?

Nur indem solche Fragen für verschiedene Arten von digitalen Objekten und in verschiedenen Kontexten beantwortet – und diese Antworten dann in konkrete Maßnahmen umgesetzt – werden, kann in der Tat von einem funktionierenden Langzeitarchivierungssystem gesprochen werden.

\*\*\*

*Dr. Max Vögler ist Referent bei der Deutschen Forschungsgemeinschaft in der Gruppe Wissenschaftliche Literaturversorgungs- und Informations-*

*systeme. Er ist Historiker und hat 2005 mit einem Thema zur Religionsgeschichte in der Habsburgmonarchie an der Columbia Universität in New York promoviert. Von 2002 bis 2005 war er Projektkoordinator bei Clio-online. E-Mail: Max.Voegler@dfg.de*

DIE EUBAM-GRUPPE UND IHRE MITWIRKUNG IN DEN  
EUROPÄISCHEN KOOPERATIONSPROJEKTEN  
MINERVA UND MICHAEL

*von Monika Hagedorn-Saupe*

*In diesem Beitrag wird die EUBAM-Gruppe vorgestellt, die als beratendes Gremium zu einer verstärkten Präsenz deutscher Kultureinrichtungen in Europa beiträgt und den Dialog zwischen den Kultureinrichtungen und europäischen Stellen fördert. Mit dieser Zielsetzung wirkt EUBAM auch mit an europäischen Kooperations- und Koordinierungsprojekten.*

\*\*\*

EUBAM steht für „Europäische Angelegenheiten für Bibliotheken, Archive, Museen und Denkmalpflege“ in Deutschland und beschreibt eine Arbeitsgruppe, die sich aus Vertretern der Kultusministerkonferenz (KMK), der Bundes- und Länderministerien, der Deutschen Forschungsgemeinschaft (DFG) sowie von Experten/innen der Sparten Bibliothek, Archiv, Museum und Denkmalpflege zusammensetzt. EUBAM wurde im Jahr 2001 ins Leben gerufen, um — wie im Leitbild (siehe Abbildung) im April 2005 formuliert — in fachlicher und strategischer Hinsicht Maßnahmen zur Verbesserung des Zugangs zu und der Erhaltung von beweglichem und unbeweglichem Kulturgut zu entwickeln und zur Koordinierung von Digitalisierungsaktivitäten im europäischen Kontext Beiträge zu leisten. EUBAM nimmt eine nationale Focus- und Transferfunktion für entsprechende Förderprogramme der EU wahr, indem sie

- die Entwicklung einer nationalen (Bund und Länder) Digitalisierungsstrategie – unter dem Gesichtspunkt des Zugangs zu kulturellen und wissenschaftlichen Inhalten für Bibliotheken, Archive, Museen und

- der Denkmalpflege – initiiert und begleitet,
- entsprechende Aktivitäten in Deutschland in moderierender Weise anregt und verbindet,
- den Aufbau nationaler, spartenübergreifender Kooperationsstrukturen fördert,
- den Informationstransfer zwischen den EU-Gremien und den oben genannten nationalen Kultursparten organisiert (unter anderem mit Internet-Angebot, Mailingliste, Informationsveranstaltungen und Erstberatung von Antragstellern),
- nationale Interessen im Vorfeld von Planungen zu EU-Förderprogrammen bündelt und an die politischen Instanzen weiter leitet sowie
- im Rahmen ihrer Mitgliedschaft in koordinierenden EU-Projekten, wie MINERVA und MINERVA-PLUS sowie MICHAEL und MICHAEL-PLUS, nationale Positionen in den entsprechenden EU-Gremien vertritt.

Vorsitzender von EUBAM ist derzeit Prof. Dr. h.c. Klaus-Dieter Lehmann, Präsident der Stiftung Preußischer Kulturbesitz in Berlin. Das Sekretariat von EUBAM wird kooperativ geführt durch die Staatsbibliothek zu Berlin Preußischer Kulturbesitz (für den Bereich Bibliotheken), das Institut für Museumskunde der Staatlichen Museen zu Berlin Preußischer Kulturbesitz (für den Bereich der Museen), das Bundesarchiv (für den Bereich der Archive) und die Oberste Denkmalschutzbehörde der Senatsverwaltung für Stadtentwicklung in Berlin (für den Bereich der Denkmalpflege).

#### **Leitbild für EUBAM**

Die interministerielle Bund-Länder-Arbeitsgruppe EUBAM (**EU**ropäische **A**ngelegenheiten für **B**ibliotheken, **A**rchive und **M**useen) ist ein Zusammenschluss von Vertretern der Kultusministerkonferenz (KMK), der Bundes- und Länderministerien, der Deutschen Forschungsgemeinschaft (DFG) sowie von Experten der Sparten Bibliothek, Archiv, Museum und Denkmalpflege. Ihr vorrangiges Ziel ist, in fachlicher und strategischer Hinsicht deren Aktivitäten und Interessen vor allem für Maßnahmen zur Verbesserung des Zugangs zu und der Erhaltung von beweglichem und

unbeweglichem Kulturgut zu entwickeln und zur Koordinierung Beiträge zu leisten. Die EUBAM nimmt damit insbesondere für Maßnahmen zur Erhaltung des wissenschaftlichen und kulturellen Erbes eine nationale Focus- und Transferfunktion für entsprechende Förderprogramme der EU wahr, indem sie

- die Entwicklung einer nationalen (Bund und Länder) Digitalisierungsstrategie – unter dem Gesichtspunkt des Zugangs zu kulturellen und wissenschaftlichen Inhalten – für Bibliotheken, Archive, Museen und der Denkmalpflege initiiert und begleitet,
- entsprechende Aktivitäten in Deutschland in moderierender Weise anregt und verbindet,
- den Aufbau nationaler, spartenübergreifender Kooperationsstrukturen fördert,
- den Informationstransfer zwischen den EU-Gremien und den oben genannten nationalen Kultursparten organisiert (unter anderem mit Internet-Angebot, Mailingliste,
- Informationsveranstaltungen und Erstberatung von Antragstellern),
- nationale Interessen im Vorfeld von Planungen zu EU-Förderprogrammen bündelt und an die politischen Instanzen weiter leitet sowie
- im Rahmen ihrer Mitgliedschaft in MINERVA und MINERVA-PLUS nationale Positionen in den entsprechenden EU-Gremien vertritt.

Die EUBAM unterstützt die interessierten Institutionen bei der Vertretung ihrer Interessen gegenüber amtlichen Stellen der Europäischen Union und unterstützt staatliche Stellen in Deutschland bei der Formulierung von Politikzielen im nationalen und europäischen Kontext sowie bei deren Implementierung. Durch Information, Beratung und Koordination leistet die EUBAM einen unverzichtbaren Beitrag zur Schaffung der für die Wettbewerbsfähigkeit notwendigen Transparenz und Motivation für die oben genannten Kultureinrichtungen in Deutschland nach innen und nach außen und unterstützt so im Sinne eines kooperativen Föderalismus die Bemühungen der Länder und des Bundes auf dem Weg des „Fit-Machens

für Europa“.<sup>1</sup>

EUBAM nimmt damit zunächst strategische Aufgaben wahr. Um diese auch effektiv umzusetzen und zum einen die für europäische Netzwerke notwendigen Stellungnahmen zu erarbeiten sowie zum anderen die Kultursparten auch intensiv zu beteiligen, wurde im Jahr 2004 die EUBAM-Arbeitsgruppe „Digitalisierung im Kulturbereich in Deutschland“ (auch als EUBAM-MINERVA-Spiegelgruppe bezeichnet) eingerichtet. Ernannt wurden sechzehn Experten/innen aus Bibliotheken, Archiven, Museen und dem Bereich der Denkmalpflege, deren Aufgabe es ist, Arbeitspapiere und Stellungnahmen zu den auf europäischer Ebene diskutierten Themen vorzubereiten und Vorschläge zu erarbeiten, wie deutsche Museen, Bibliotheken, Archive und Einrichtungen der Denkmalpflege im europäischen Kontext sichtbar gemacht werden können. Auf Bitten von EUBAM hat die Arbeitsgruppe ein Konzept für ein Internet-Portal „Kulturerbe-digital“ entwickelt. Mit dem Aufbau dieses Internet-Portals soll die deutsche und internationale Öffentlichkeit in umfassendem Maße und gebündelt über die zahlreichen deutschen Aktivitäten im Bereich der Digitalisierung von Kulturgut informiert werden. Das Portal soll die vielfältigen vorhandenen Ressourcen miteinander vernetzen. Durch die Zusammenarbeit mit dem „Projektträger Neue Medien in der Bildung + Fachinformation (PT-NMB+F) im DLR (Deutsches Zentrum für Luft- und Raumfahrt e.V.)“ ist auch die Internet-Plattform **DL-Forum**<sup>2</sup> beteiligt, die Informationen zu Projekten, Förderprogrammen, Förderbekanntmachungen, Veranstaltungen und Anlaufstellen im Umfeld des Förderschwerpunkts „Digitale Bibliothek“ insbesondere des Bundesministeriums für Bildung und Forschung (BMBF) und der Deutschen Forschungsgemeinschaft (DFG) zusammenführt.

In Kooperation mit dem DL-Forum erarbeiten die Sekretariate gemeinsam mit der EUBAM-Arbeitsgruppe „Digitalisierung im Kulturbereich

1 Vorsitzender der EUBAM ist derzeit Prof. Dr. h.c. Klaus-Dieter Lehmann, Präsident der Stiftung Preußischer Kulturbesitz, Berlin. Website: <<http://www.eubam.de>> (Am 19.04.2005 verabschiedete Fassung). Alle Links wurden zuletzt am 06.11.2006 überprüft.

2 Vgl. <<http://www.dl-forum.de>>.

reich in Deutschland“ eine Übersicht über Digitalisierungsprojekte im Kulturbereich. Diese Informationen, zunächst als Liste vorgelegt und auf der Homepage von EUBAM<sup>3</sup> zum Herunterladen bereitgestellt, werden nun in eine Datenbank eingepflegt, so dass sie in Zukunft besser recherchierbar allen Interessierten zur Verfügung stehen.

### MINERVA

Die EUBAM-Arbeitsgruppe „Digitalisierung im Kulturbereich in Deutschland“ wird auch als „EUBAM-MINERVA-Spiegelgruppe“ bezeichnet, da sie die Schnittstelle bilden soll zwischen den deutschen und europäischen Diskussionen zum Themenkomplex „Digitalisierung im Kulturbereich“. MINERVA (Ministerial NETwork for Valorising Activities in digitisation) begann als im Rahmen des IST-Programms gefördertes Projekt (Laufzeit: 2002-2005) von überwiegend Kultusministerien mit dem Ziel, eine europäische Plattform für Digitalisierungsaktivitäten im kulturellen Bereich zu schaffen, die nationalen Digitalisierungsstrategien und -aktivitäten stärker aufeinander abzustimmen und gemeinsam mit der Europäischen Kommission die im eEurope-Aktionsplan festgelegten Ziele zu erreichen.<sup>4</sup>

Die Europäische Kommission hatte im Dezember 1999 die Initiative eEurope gestartet, um Europa „ans Netz zu bringen“. Hierfür wurde ein eEurope-Aktionsplan erstellt. Im März 2000 in Lissabon hat die Kommission auf dem Treffen des Europäischen Rates einen Bericht vorgelegt. Am 19./20. Juni 2000 in Feira (Portugal) hat der Europäische Rat dem eEurope-Aktionsplan 2002 zugestimmt. Dieser Aktionsplan umfasste auch den Bereich „Europäische digitale Inhalte für globale Netze“, wobei explizit auch der kulturelle Bereich genannt wurde. Dem eEurope-Aktionsplan 2002 folgte der eEurope-Aktionsplan 2005, diesem wiederum die Initiative i2010, die von der Kommission im Juni 2005 verabschiedet wurde.<sup>5</sup> In i2010 wird formuliert, dass eine Informationsgesellschaft gefördert werden soll, die **alle Menschen einbezieht**. Um diese zu erreichen, sollen eine

3 Vgl. <<http://www.eubam.de>>.

4 Vgl. <<http://www.minervaeurope.org>>.

5 Vgl. <[http://ec.europa.eu/information\\_society/eeurope/i2010/index\\_en.htm](http://ec.europa.eu/information_society/eeurope/i2010/index_en.htm)>.

Reihe von Initiativen angeregt und umgesetzt werden. Eine davon ist der Aufbau von digitalen Bibliotheken, über die alle Zugang zu Multimedia und zur multilingualen europäischen Kultur haben sollen.

Für Deutschland erhob sich die Frage, wer Partner in MINERVA werden könnte. Bedingt durch die Kulturhoheit der Länder, liegt die Hauptverantwortung für Kultur bei den sechzehn Kultusministerien der Länder, daneben nimmt der/die Bundesbeauftragte für die Kultur und die Medien, angesiedelt am Bundeskanzleramt, eine Reihe von Aufgaben wahr. Aufgrund dieser verteilten Verantwortlichkeiten für Kultur in Deutschland wurde die Stiftung Preußischer Kulturbesitz für EUBAM Partner in MINERVA. Im Februar 2004 wurde MINERVA durch MINERVA Plus erweitert. MINERVA Plus ist die Erweiterung und Fortführung des bis 2005 befristeten Projektes MINERVA. Das Projekt umfasst die Kooperation zwischen den Ländern Österreich, Belgien, Tschechien, Dänemark, Estland, Finnland, Frankreich, Deutschland, Griechenland, Großbritannien, Ungarn, Niederlande, Irland, Italien, Malta, Polen, Portugal, Spanien, Schweden, und Slowenien (Israel und Russland sind als assoziierte Partner beteiligt).

Das MINERVA-Projekt hatte zum Ziel, zum einen koordinierende Maßnahmen zur Entwicklung abgestimmter Digitalisierungsstrategien im Kulturbereich zu unterstützen und zu befördern; zum anderen sollten europäische Arbeitsgruppen abgestimmte Empfehlungen und Arbeitsmaterialien erstellen, die die Digitalisierung im Kulturbereich und das Zugänglichmachen der Digitalisate unterstützen. Eine Arbeitsgruppe mit dem Titel „Identification of user needs, content and quality framework for common access points“ behandelt die Fragen der Zugänglichmachung der Daten im Internet. Gemeinsam wurde ein Handbuch entwickelt („Handbook für Quality in Cultural Websites: Improving Quality for Citizens“), in dem Rahmenbedingungen und Grundprinzipien formuliert wurden, die von Relevanz sind, wenn man eine qualitativ hochwertige Website aufbauen möchte. Ausgehend von diesen allgemeinen Grundsätzen wurden zehn Anforderungen an eine gute Webseite formuliert:

„Eine qualitativ hochwertige Webseite erfüllt folgende Kriterien:

- **Transparenz:** sie soll deutlich strukturiert sein und den Zweck sowie

die Verantwortlichen für die Seite benennen;

- **Effektivität:** sie soll ihre Inhalte sinnvoll auswählen, digitalisieren, präsentieren und validieren, um eine größtmögliche Nutzbarkeit zu gewährleisten;
- **Pflege:** sie soll Richtlinien für eine Qualitätssicherung beachten und damit sicher stellen, dass die Webseite gepflegt und in angemessenem Zeitraum aktualisiert wird;
- **Zugänglichkeit:** sie soll für alle Benutzer zugänglich sein, unabhängig von der verwendeten Technologie und von persönlichen Behinderungen; dies schließt Navigation, Inhalt und alle interaktiven Elemente ein;
- **Benutzerorientierung:** sie soll die Nutzerbedürfnisse berücksichtigen, Relevanz und leichte Bedienbarkeit gewährleisten sowie Evaluationsergebnisse und Feedback einbeziehen;
- **Interaktivität:** sie soll den Nutzern die Möglichkeit bieten, mit den Verantwortlichen Kontakt aufzunehmen und eine zweckdienliche Antwort zu erhalten. Sie soll – wo es angemessen ist – die Nutzenden ermutigen, Anfragen zu stellen und Informationen gemeinsam zu nutzen sowie Diskussionen mit und unter den Nutzern anregen;
- **Mehrsprachigkeit:** sie soll sich der Wichtigkeit der Mehrsprachigkeit bewusst sein, indem sie als Mindeststandard einen Zugang in mehr als einer Sprache anbietet;
- **Interoperabilität:** sie soll in kulturelle Netzwerke integriert sein und es den Nutzenden ermöglichen, Inhalte und Services, die ihren Bedürfnissen entsprechen, leicht aufzufinden;
- **verantwortliches Management:** sie soll den rechtlichen Bestimmungen wie Urheberrecht und Datenschutz entsprechen. Die Bedingungen der Nutzung der Webseite und ihrer Inhalte müssen deutlich gemacht werden;
- **dauerhafte Bewahrung:** sie soll solche Richtlinien und Standards übernehmen, die sicherstellen, dass die Webseite und ihr Inhalt langfristig erhalten werden können.<sup>6</sup>

6 Vgl. <<http://www.minervaeurope.org>>.

Für jeden dieser Punkte wurde eine Checkliste aufgestellt sowie einige Fragen formuliert, die die Verantwortlichen für die Webpräsentation in Kultureinrichtungen darin unterstützen sollen, die eigene Webseite auf Qualität zu prüfen und nutzergerechte Webseiten aufzubauen. Diese zehn Prinzipien und die ihnen zugeordneten Checklisten und Fragen wurden in mehreren europäischen Sprachen publiziert (weitere sind in Vorbereitung). Während große Kultureinrichtungen oft bereits einen ausgebauten Internetauftritt haben, ist dies für kleinere Museen, Bibliotheken und Archive oft noch nicht der Fall. Um diese Einrichtungen beim Aufbau einer eigenen Webseite zu unterstützen, wurde ein weiteres Instrument entwickelt: „Museo & Web“. Hierbei handelt es sich um eine webbasierte Anleitung für den Aufbau einer Webseite, die den zehn Qualitätskriterien genügt. Eine weitere Arbeitsgruppe erarbeitete ein Handbuch, das die Kultureinrichtungen bei der Durchführung von Digitalisierungsprojekten unterstützen soll, das in zehn europäischen Sprachen verbreitet wird. Der deutsche Titel lautet: „Good Practice – Handbuch für Digitalisierungsprojekte“. Eine Arbeitsgruppe unter britischer Leitung erarbeitete technische Richtlinien, die bei Digitalisierung berücksichtigt werden sollten. Auch diese Empfehlung liegt in vier Sprachen (Deutsch, Englisch, Französisch und Griechisch) vor. Alle diese Publikationen stehen auch auf der Homepage von MINERVA/MINERVAPLUS<sup>7</sup> zur Verfügung und viele Interessierte haben inzwischen davon Gebrauch gemacht. Ende Januar 2006 ist die finanzielle Förderung der EU von MINERVA/MINERVA-PLUS ausgelaufen. Insgesamt haben sich alle beteiligten Partner weit über das vertraglich festgelegte engagiert und führen auch nach Auslaufen der Projektförderung gemeinsame Aktivitäten weiter.

Aus den gemeinsamen Aktivitäten zur Förderung von Digitalisierung im Kulturbereich heraus entstanden auch die Überlegungen zum Aufbau eines europäischen Portals, in dem die im Kulturbereich vorhandenen digitalen Sammlungen aus den verschiedenen europäischen Ländern nachgewiesen werden. Die Kultusministerien Frankreichs und Italiens und die dem britischen Kultusministerium nachgeordnete Einrichtung MLA

7 Vgl. <<http://www.minervaeurope.org>>.

(Museums, Libraries and Archives Council), die in ihren Ländern Digitalisierungsvorhaben und die Veröffentlichung von digitalen Daten in großem Umfang fördern, beantragten eine Förderung bei der EU, um die nationalen Aktivitäten europäisch zu vernetzen. Dies ist das Projekt MICHAEL.

### MICHAEL

Im Rahmen des eTEN-Programms wird inzwischen das Vorhaben MICHAEL (Multilingual Inventory of Cultural Heritage in Europe)<sup>8</sup> gefördert. Zielsetzung des Projektes ist es, europäisches kulturelles Erbe weltweit zugänglich zu machen. MICHAEL ist darauf ausgerichtet, die vielfältigen europäischen digitalen Sammlungen von Museen, Bibliotheken und Archiven zu vernetzen und diese über ein zentrales Zugangportal zu erschließen. MICHAEL soll damit die Recherche schwer auffindbarer Informationen zu digitalen Sammlungen von Europas Kultureinrichtungen erleichtern. Unter Nutzung der bereits vom französischen Kultusministerium genutzten Open Source-Technologie für den Nachweis von digitalen Sammlungen im Kulturbereich in Frankreich<sup>9</sup> wird ein mehrsprachiges Portal aufgebaut. Das für die Präsentation der digitalen Sammlungen formulierte Datenmodell basiert auf den Empfehlungen von MINERVA zur Interoperabilität und orientiert sich an den W3C-Empfehlungen für die Beschreibung von digitalen Sammlungen. Die Software-Plattform umfasst zwei Module: Ein Produktionsmodul, mit dessen Hilfe webbasierte Inhalte in das jeweilige Portal eingepflegt werden können und ein Publikationsmodul. Letzteres bietet ein intuitives, einfach nutzbares Interface, damit interessierte Nutzer/innen mit ihrem Webbrowser digitale kulturelle Inhalte leicht auffinden können. Mit Hilfe dieses Publikationsmoduls werden nationale MICHAEL-Portale aufgebaut, die über eine OAI-PMH-Schnittstelle Daten an den europäischen Server liefern.

Großbritannien, Italien und Frankreich haben bereits einen solchen nationalen Server aufgesetzt. Inzwischen sind auch diesem Vorhaben elf weitere Länder (Belgien, Deutschland, Finnland, Griechenland, Malta,

<sup>8</sup> Vgl. <<http://www.michael-culture.org>>.

<sup>9</sup> Vgl. <<http://www.culture.fr>>.

Niederlande, Polen, Schweden, Spanien, Tschechische Republik, Ungarn) beigetreten (Projekt MICHAEL-PLUS), die bis 2007 weitere nationale Server einrichten werden. Eine erste Version des europäischen Servers wird voraussichtlich im November 2006 online sein. Eine besondere Herausforderung für den Aufbau des europäischen Servers ist die Sprachenvielfalt in Europa. Das Portal soll den interessierten Nutzern/innen in ihrer eigenen Sprache zur Verfügung stehen, die Navigation und allgemeine Beschreibungen müssen in allen Sprachen der beteiligten Partner vorliegen. Damit die Nutzer/innen über alle Sammlungen in ihrer eigenen Sprache recherchieren können, wird zunächst auf ausgewählte Begriffe des UNESCO-Thesaurus aufbauend eine Liste von Begriffen mit ihren jeweiligen Entsprechungen in den Sprachen der Partner im Portal implementiert, über die die Suche erfolgen wird. In der Regel sind auch in MICHAEL/MICHAEL-PLUS wie auch in MINERVA/MINERVA-PLUS die jeweiligen Kultusministerien Projektpartner. Für Deutschland hat EUBAM wegen der verteilten Verantwortlichkeiten für Kultur einen anderen Weg vorgeschlagen und es konnten sieben große Kultureinrichtungen als Partner für das Projekt gewonnen werden: das Bundesarchiv und das Landesarchiv in Baden-Württemberg, Die Deutsche Bibliothek und die Bayerische Staatsbibliothek, das Deutsche Museum in München und das Naturmuseum Senckenberg in Frankfurt am Main sowie die Stiftung Preußischer Kulturbesitz, die mit der Staatsbibliothek zu Berlin, den Staatlichen Museen zu Berlin und dem Geheimen Staatsarchiv alle drei Kultursparten umfasst. Diese sieben Partner werden gemeinsam das deutsche Portal aufbauen und dafür werben, dass möglichst viele deutsche Kultureinrichtungen Informationen über ihre digitalen Sammlungen für das deutsche und europäische Portal bereitstellen.

Auch EUBAM und die EUBAM-MINERVA-Spiegelgruppe unterstützen den Auf- und Ausbau des MICHAEL-Portals, weil damit die Vielfalt und der Reichtum der Kultur in Deutschland besser sichtbar werden.

Auf europäischer Ebene werden derzeit große Anstrengungen unternommen, digitalisiertes bzw. digitales Kulturerbe möglichst schnell und in möglichst großem Umfang einem weltweiten Publi-

kum zugänglich zu machen. Man spricht von der Europäischen Digitalen Bibliothek, die Informationen zu Beständen aus allen Kultursparten umfassen soll. Um mit dieser rasanten Entwicklung Schritt halten zu können, muss schnell reagiert werden können. Hier ist EUBAM als koordinierendes Gremium eine geeignete Instanz.

\*\*\*

*Monika Hagedorn-Saupe ist Leiterin des Referats „Besucherbezogene Museumsforschung und Kulturstatistik“ am Institut für Museumskunde der Staatlichen Museen zu Berlin Preußischer Kulturbesitz und stellvertretende Leiterin des Instituts Lehraufträge in den Bereichen Besucherforschung, Museumspädagogik und Internationale Medieninformatik an der Fachhochschule für Technik und Wirtschaft in Berlin. E-Mail: m.hagedorn@smb.spk-berlin.de*

DIE DIGITALISIERUNGSPROJEKTE DER  
UNIVERSITÄTSBIBLIOTHEKEN IN FRANKREICH –  
PERSEE UND ANDERE ANGEBOTE

*von Frédéric Blin und Valérie Néouze*

*Digitalisierungsprogramme werden in den französischen Bibliotheken seit ungefähr 20 Jahren geleitet. Heutzutage führen die meisten Einrichtungen, wie etwa öffentliche Bibliotheken (Ministère de la Culture) oder akademische Bibliotheken (Ministère de l'Education nationale, de l'enseignement supérieur et de la recherche = MENESR, Ministerium für Bildung und Forschung), Digitalisierungsprojekte ihrer Sammlungen durch. Gallica, die digitale Bibliothek der Bibliothèque nationale de France mit ihren circa 70.000 Werken im Volltext und 80.000 Bildern, ist das bekannteste französische Beispiel solcher Programme auf internationaler Ebene. Darüber hinaus werden zurzeit weitere wichtige Projekte in den akademischen Bibliotheken entwickelt. Diese Vorhaben werden vom MENESR finanziell unterstützt. Dabei entstehen auch nationale Programme, wie PERSEE, das französische Portal für retrospektive Sammlungen wissenschaftlicher Zeitschriften in den Geistes- und Sozialwissenschaften in französischer Sprache.*

\*\*\*

**Die Digitalisierungsprojekte akademischer Bibliotheken in Frankreich**

Nationalen und internationalen Beispielen folgend, sind in den letzten Jahren zahlreiche lokale Digitalisierungsprojekte in den französischen Universitätsbibliotheken initiiert worden. Gleichzeitig reagieren sie damit ebenfalls auf eine neue Realität des wissenschaftlichen Veröffentlichungswesens. Diese Projekte, welche die auf französischem Niveau angestellten

Überlegungen über die Schaffung einer europäischen digitalen Bibliothek vorangetrieben haben<sup>1</sup>, deuten die Grundlagen eines digitalen dokumentarischen Netzes an, das jenes der CADIST-Bibliotheken (*Centres d'Acquisition et de Diffusion de l'Information Scientifique et Technique* – Zentren des Erwerbs und der Verbreitung der wissenschaftlichen und technischen Information<sup>2</sup>) vervollständigt.

Unter diesem Gesichtspunkt haben einige Bibliotheken seit mehreren Jahren Portale zu digitalisierten Dokumenten, von nationalem oder sogar internationalem Interesse, entwickelt. Unter den repräsentativsten Verwirklichungen in diesem Bereich kann man folgende erwähnen:

- *Medic@*<sup>3</sup>, die digitale Bibliothek für Medizin der *Bibliothèque interuniversitaire de Médecine de Paris* (BIUM), präsentiert aus ihren Sammlungen mehr als 3.200 Werke im Volltext, sowie eine Datenbank von Bildern und Arztporträts von fast 6.000 Dokumenten. Die BIUM beteiligt sich darüber hinaus mit ihren Sammlungen auch an anderen Digitalisierungsprojekten, so etwa im Portal *Criminocorpus*, welches zum Beispiel die Zeitschrift *Archive de l'anthropologie criminelle* von 1886 bis 1914 im Volltext zugänglich macht.<sup>4</sup> Damit hat sich das Internetprojekt der BIUM mit seinem föderalen Charakter letztendlich zu einer maßgeblichen, virtuellen Medizinbibliothek entwickelt.<sup>5</sup>
- Das *Conservatoire numérique des Arts et Métiers* (CNUM<sup>6</sup>) ist eine virtuelle Bibliothek, die sich der Wissenschafts- und Technikgeschichte widmet. Sie besteht aus den Sammlungen des *Conservatoire natio-*

1 Jeanneney, Jean-Noël, *Quand Google défie l'Europe*, Paris 2005.

2 Diese CADIST-Bibliotheken sind, wie die Sondersammelgebietsbibliotheken in Deutschland, spezialisierte Forschungsbibliotheken. Sie wenden sich hauptsächlich an die Gesamtheit der französischen Wissenschaftler/innen in ihren spezifischen Bereichen.

3 *MEDIC@*, vgl. <<http://www.bium.univ-paris5.fr/histmed/medica.htm>> (Alle Links wurden zuletzt überprüft am 20.09.2006).

4 *Criminocorpus*, vgl. <<http://www.hstl.crhst.cnrs.fr/criminocorpus>>.

5 Unter den angebotenen Diensten findet man Verbindungen in Richtung Bücher und elektronischer Zeitschriften, spezialisierter Internet-Adressen, virtueller Ausstellungen, Bilddatenbanken, Namensverzeichnissen und Zeitplänen usw.

6 *Conservatoire numérique des arts et métiers*, vgl. <<http://cnum.cnam.fr/>>.

*nal des arts et métiers* (CNAM<sup>7</sup>). Das Portal, das mehr als 400 digitalisierte Werke vom 16. bis zum 20. Jahrhundert (circa 120.000 Seiten), darunter auch die nationalen Ausstellungen und Weltausstellungen des 19. Jahrhunderts, anbietet, registriert etwa 25.000 Besucher/innen im Monat.

- Die Manuskripte der Rheinmystik der *Bibliothèque nationale et universitaire de Strasbourg*: Es handelt sich um 25 religiöse Bücher des 14. und 15. Jahrhunderts. Diese wurden als Bild und nicht als Text digitalisiert.<sup>8</sup>
- Die Online-Ausgaben der *Ecole nationale des chartes*: Eine große Bildungs- und Forschungseinrichtung, die auf historische Wissenschaften spezialisiert ist. Sie verfolgt eine aktive Politik der Digitalisierung und Publikation historischer Quellen, wie etwa Chroniken, Chartulari, Prägestempel, Manuskripte usw.<sup>9</sup> Diese Dokumente werden mit Kommentaren und kritischen Anmerkungen von den Lehrenden und Studierenden der Schule ergänzt. Die Bibliothek begleitet dieses Programm, indem sie ihre Sammlungen zur Verfügung stellt.
- Der Nachlass des Wissenschaftlers Jean-Martin Charcot wurde gemeinsam mit dem Dokumentationsdienst der Universität Paris VI. „Pierre und Marie Curie“ digitalisiert.<sup>10</sup> Benutzer/innen können dort Unterrichtsmanuskripte, Thesen, Zeitschriften und seltene Werke sowie Alben der Assistenzärzte/innen des Krankenhauses *de la Salpêtrière* finden.
- Zahlreiche Universitäten haben ebenfalls ihre Dissertationsbestände digitalisiert. In einigen Fällen stellen sie sogar ihre wissenschaftlichen Gutachten zur Verfügung.

Weitere Projekte mit einer national koordinierten Digitalisierungspolitik sind zum Beispiel die Projekte der Bibliothek Cujas in Paris, französische

7 *Conservatoire national des arts et métiers*, vgl. <<http://www.cnam.fr/>>.

8 Digitalisate der *Bibliothèque nationale et universitaire de Strasbourg*, vgl. <<http://www-bnus.u-strasbg.fr/BNU/FR/>>.

9 Die *Editions en ligne de l'Ecole nationale des chartes*, vgl. <<http://elec.enc.sorbonne.fr/>>.

10 Charcot Nachlass: <<http://jubil.upmc.fr/repons/portal/bookmark?MainTab=CMS>ShowChannelDoc&ShowDocChannel=presentation/SCDM/charcot/historique>>.

Hauptbibliothek für Rechtswissenschaften<sup>11</sup>, und die *Bibliothèque de documentation internationale contemporaine* (BDIC), Hauptbibliothek für Zeitgeschichte<sup>12</sup> mit ihren alten und wertvollen Sammlungen. Da diese Projekte noch sehr jung sind, konnten bisher nur eine geringe Zahl von Digitalisaten, beispielsweise historische, rechtswissenschaftliche Werke der Cujas-Bibliothek, online zur Verfügung gestellt werden. Alle diese Projekte wurden von den lokalen Einrichtungen initiiert und unter anderem vom Ministerium für Bildung und Forschung gefördert. In der Tat werden die erwähnten Vorhaben – entweder im Rahmen vierjähriger, staatlicher Verträge, oder im Rahmen außergewöhnlicher Aktionen – durch die Ausschüsse des Ministeriums (Unterabteilung der Bibliotheken) geprüft, die dadurch den Bibliotheken helfen können, ihre wissenschaftlichen und technischen Projekte zu entwickeln. Die Bewilligung einer finanziellen Unterstützung erfolgt mittels mehrerer Kriterien:

- Wissenschaftliche Qualität der Inhalte sowie des kritischen Apparats, das die digitalisierten Dokumente begleitet,
- die Wahl der technischen Lösungsansätze (Qualität der Digitalisierung, Textformat bezüglich des Bildformates, Integration in die Informationssysteme der Universitäten),
- Zugang zum Produkt der Digitalisierung (Suchmaschinen, Sudoc-Portal<sup>13</sup>, OAI-Harvesting...). Die Frage des Zugangs, bisher zu wenig berücksichtigt, gilt nun als essentiell und wird deshalb besonders vom Ministerium geprüft.

Durch diese Begutachtungs- und Unterstützungsrolle will das Ministerium die Koordinierungspolitik der Digitalisierungsprogramme auf eine nationale Ebene stellen. Um besser zu identifizieren, welche Sammlungen schon digitalisiert worden sind oder wo bereits Digitalisierungsprogramme angelaufen sind, aber auch um eine Annäherung lokaler Projekte zu för-

11 Bibliothek Cujas, vgl. <<http://biu-cujas.univ-paris1.fr/>>.

12 Bibliothèque de documentation internationale contemporaine (BDIC), vgl. <<http://www.bdic.fr/index.php3>>.

13 SUDOC Portal, vgl. <<http://www.portail-sudoc.abes.fr/>>. Das Sudoc-Portal ist das fördernde Portal zu Online-Ressourcen der französischen Hochschulen und Universitäten.

dern, hat das Ministerium Ende 2005 beschlossen, eine Datenbank zu den schon digitalisierten Beständen in den französischen Hochschulen einzurichten. Um dieses Ziel zu erreichen, haben sich das Ministerium für Bildung und das Ministerium für Kultur zu einer Zusammenarbeit entschlossen. Das Ministerium für Kultur hat im Rahmen des europäischen Programms MICHAEL (Multicultural Inventory of Cultural Heritage in Europe<sup>14</sup>) sein eigenes Portal zu den digitalisierten kulturellen Beständen seiner Institutionen, wie öffentliche Bibliotheken, Archive und Museen, gebaut. Das MENESR hat also das von MICHAEL benutzte Datenmodell überprüft, um es an die Anforderungen der Bestände der Hochschulen, mit Hinblick auf eine technische Kompatibilität mit dem Portal MICHAEL, anzupassen. Das besagte Portal zu den digitalisierten Beständen der französischen Hochschulen soll Ende des Jahres 2006 zugänglich gemacht werden.

Parallel zu dieser Unterstützung lokaler Projekte kann das Ministerium ebenfalls die Initiative ergreifen, nationale Projekte einzuführen. In den letzten Jahren sind zwei Hauptprojekte entstanden; Liber Floridus, eine Datenbank von Buchmalereien mittelalterlicher Manuskripte<sup>15</sup>, die im Besitz der Hochschulen sind, und PERSEE<sup>16</sup>, das nun etwas ausführlicher vorgestellt werden soll.

## PERSEE

### *Der Kontext*

Das Programm PERSEE erfolgt in einem schwierigen wirtschaftlichen Kontext für das Verlagswesen in den Geistes- und Sozialwissenschaften. Während die Forscher/innen in den Naturwissenschaften im Zeitschriftenbereich – Hauptvektor der wissenschaftlichen Information – seit langem die neuen Technologien nutzen, um Methoden für die elektronische Erfassung und Verbreitung der wissenschaftlichen Ergebnisse zu entwickeln und dadurch den schnellen Umlauf, die internationale Verbreitung und die

14 MICHAEL Portal, vgl. <[http://www.michael-culture.org/index\\_f.html](http://www.michael-culture.org/index_f.html)>.

15 Liber Floridus, vgl. <<http://liberfloridus.cines.fr/>>.

16 PERSEE, vgl. <<http://www.persee.fr/>>.

optimale Nutzung dieser Ergebnisse für ihre Gemeinschaft favorisieren, bleibt der Bereich der Zeitschriften in den Geistes- und Sozialwissenschaften dagegen aus mehreren Gründen noch abseits dieser Entwicklung.

Charakterisiert wird der geistes- und sozialwissenschaftliche Bereich durch eine Aufsplitterung der Akteure, unter denen die Hochschulen und die gelehrten Gesellschaften einen ausschlaggebenden Platz einnehmen. Die zahlreichen Zeitschriften befinden sich im Allgemeinen in einer sehr heiklen wirtschaftlichen Lage, die mit zwei Hauptfaktoren zusammenhängt.

- Einerseits leiden sie unter einer schwachen internationalen Sichtbarkeit, was hauptsächlich auf eine vorrangig französisch sprechende Leserschaft zurückzuführen ist, aber auch, weil diese Zeitschriften auf einem Modell der Herstellung und Verbreitung beruhen, das auf dem Papier begründet wurde, und welches sich infolgedessen stark von den neuen Praktiken der elektronischen Herausgabe von Forschungsergebnissen unterscheidet.
- Andererseits sind die Redaktionen, die einen anderen Umgang mit Informatiktechnologien pflegen als die naturwissenschaftlichen Publikationsorgane, an eine Kultur des Verlagswesens gebunden.

Trotzdem stellen die Zeitschriften in den Geistes- und Sozialwissenschaften, insbesondere in Frankreich, ein wissenschaftliches Kulturgut sehr hoher Qualität dar. Einer technischen Angleichung sehen sowohl die Forscher/innen als Erzeuger/innen und Verbraucher/innen dieser Information, als auch Redakteure/innen der Zeitschriften mit großer Ungeduld entgegen, da dies die nationale und internationale Sichtbarkeit ihrer Produktion fördern und die Dauerhaftigkeit einiger extrem spezialisierter Titel gewährleisten könnte.

Der Wert dieser Zeitschriften und der potentielle wirtschaftliche Mehrwert, den sie darstellen, sind den angelsächsischen Hauptakteuren nicht verborgen geblieben. Seit mehreren Jahren unternehmen sie eine systematische Akquisition französischer Publikationen aus dem Bereich der Geistes- und Sozialwissenschaften, mit den anspruchsvollsten Titeln beginnend. Sie haben ihren französischen Kollegen/innen vorgeschlagen, die digitalisierten Zeitschriften in ihre eigenen virtuellen und kostenpflich-

tigen Sammlungsportale zu integrieren. Die Angebote dieser Datenverteiler bzw. Akteure sind hinsichtlich der Strategien zur Verbreitung der französischen Wissenschaftsforschung sehr aufschlussreich. In der Tat hätte im Falle mangelnder Alternativen die Wahl dieser kommerziellen Logik in großem Ausmaße verhängnisvolle Folgen gehabt:

- Höhere Kosten, die die wissenschaftliche Gemeinschaft hätte übernehmen müssen, um Zugang zu diesen Zeitschriften zu haben. Das heißt paradoxerweise, dass sie dafür bezahlen müsste, ihre eigenen Forschungsergebnisse zu befragen, obwohl diese schon weitgehend durch öffentliche Mittel subventioniert wurden.
- Das Risiko für die Zeitschriften in den Geistes- und Sozialwissenschaften, einer fast monopolistischen elektronischen Verbreitung zu unterliegen, besteht und kann im Bereich der Naturwissenschaften bereits beobachtet werden.
- Eine weitere Folge wäre eine zusätzliche Schwächung eines Teiles der wissenschaftlichen Produktion, die sehr spezialisiert ist und in Zeitschriften veröffentlicht wird, die in einer strikten Handelslogik für nicht rentabel gehalten würden. Es könnte sogar das Ende einiger Titel wegen einer erzwungenen Koordinierung bedeuten, welche die Konkurrenz zwischen den eigenen Titeln eines Datenverteilers vermeiden soll.

In den Geistes- und Sozialwissenschaften scheint die Mehrzahl der Herausgeber/innen unter bestimmten Bedingungen nicht gegen einen freien Zugang zu den digitalisierten Auflagen der Zeitschriften zu sein. Die Bewilligung einer öffentlichen finanziellen Beihilfe für die retrospektive Digitalisierung der Sammlungen als Gegenleistung zu ihrer kostenlosen Bereitstellung im Netz könnte sich ihrer Meinung nach wie folgt gestalten:

- Ein kostenloses Onlinestellen der Zeitschriften, im Durchschnitt drei bis fünf Jahre nach ihrer ersten Veröffentlichung, bedeutet in der Regel keinen Verdienstausschlag, da wenig Artikel oder thematische Faszikel eine langfristige wirtschaftliche Rentabilität haben.
- Die Verbreitung kostenloser Artikel auf dem Internet stellt ein wirksames Abrufprodukt dar, welches eine neue Leserschaft anziehen könnte, die möglicherweise für ein Abonnement der laufenden Veröffentlichungen in Frage käme.

- Die Handelsverbreitung der letzten Lieferungen (mit Varianten bezüglich der Dauer der *Moving Wall*, die im Durchschnitt für einen Zeitraum von drei bis fünf Jahren zu bewahren ist, um den Verkauf der Zeitschriften zu gewährleisten) soll aufrechterhalten werden.
- Die gedruckte Ausgabe, parallel oder in Anlehnung an die elektronische Version ihrer Veröffentlichungen, soll beibehalten werden.

#### *Die Richtlinien von PERSEE*

Aus diesem Protokoll heraus ist das Projekt PERSEE als öffentliches Portal für die elektronische Verbreitung der retrospektiven Sammlungen französischsprachiger Zeitschriften der Geistes- und Sozialwissenschaften entstanden. Es gibt drei Hauptzielsetzungen:

- dieses wissenschaftliche Kulturgut durch seine Verbreitung im Internet zu bewahren und zu valorisieren,
- an der Wirkung der Forschung in französischer Sprache durch die Verbreitung eines ausgedehnten Zeitschriftencorpus teilzunehmen, um dieses in das internationale wissenschaftliche Angebot zu integrieren und so eine linguistische Isolierung zu vermeiden,
- durch die Suchmethoden, die nach der Digitalisierung ermöglicht werden, erweiterte Arbeits- und Nutzungsmöglichkeiten dieser Zeitschriften anzubieten, die jetzt schon in anderen Bereichen üblich geworden sind (automatisierte Recherche in ausgedehnten Sammlungen, Downloads, Anmerkungen usw.).

Unter Leitung der *Direction de l'enseignement supérieur* hat das Projekt PERSEE eine weitgehende Zusammenarbeit mit den anderen Institutionen (unter anderem mit der Direktion für Forschung und Direktion für Technologie) erlaubt, die das wissenschaftliche Veröffentlichungswesen aktiv unterstützen. Drei allgemeine Konzepte sind von diesen verschiedenen Partnern im Rahmen von konkreten Aktionen entwickelt worden:

Ein pragmatisches Konzept:

- Die Notwendigkeit eines kurzfristigen Aktionsplans: In einem wettbewerbsfähigen, internationalen, wissenschaftlichen Kontext und in einer Situation, in der in Frankreich noch keine Alternative vorgeschlagen wurde, um die Veröffentlichung der Zeitschriften online zu erlauben,

- haben einige Herausgeber bilaterale Verhandlungen mit den privaten Handelskonsortien begonnen. Andere haben die Idee eines Verbands bevorzugt, um ein französischsprachiges Ensemble zu schaffen – innerhalb eines hauptsächlich englischsprachigen Angebots.
- Die Wahl eines auf sieben Zeitschriften<sup>17</sup> beschränkten Corpus, die den Anfangskern des Portals bilden und die schnelle Entwicklung eines Prototyps erlauben sollen.
- Die Begrenzung auf die retrospektive Digitalisierung dieser Zeitschriften, im Rahmen des derzeitigen Kompetenzfeldes der *Direction de l'enseignement supérieur*.

Ein wissenschaftliches Konzept:

- Eine enge und ständige Kooperation mit den Forschern/innen und den Zeitschriftenredaktionen, um das Profil und die Funktionen des Portals ihren spezifischen Bedürfnissen anzupassen.
- Die Festlegung einer Auswahl an Zeitschriften verschiedener Disziplinen der französischen wissenschaftlichen Forschung.

Ein offenes Konzept:

- Im Sinne des „öffentlichen Dienstes“: ein frei und kostenlos zugängliches Portal in einer nicht exklusiven Verbreitungslogik.
- Eine Wahl offener technologischer Optionen: Informatikentwicklungen in Open Source, Wahl von international anerkannten Standards und Normen, um die notwendige Kompatibilität mit ähnlichen oder ergänzenden Initiativen im Bereich der retrospektiven Digitalisierung oder des elektronischen Publizierens zu garantieren.
- Eine internationale Öffnung, die eine aktive Partnerschaft mit ERUDIT, einem Projekt aus Quebec<sup>18</sup>, bevorzugt, dessen politische Zielsetzungen, wissenschaftliche und technologische Optionen, sowie linguistischer und disziplinarischer Umkreis jenen ähnlich sind, die soeben dargelegt wurden.

17 Die *Annales*, die *Bibliothèque de l'école des chartes*, *L'Homme*, *Matériaux pour l'histoire de notre temps*, die *Revue de l'art*, die *Revue économique*, die *Revue française de science politique*.

18 ERUDIT, vgl. <<http://www.erudit.org>>.

### Die Schaffung des Konsortiums PERSEE

Nachdem diese drei allgemeinen Grundsätze und die entsprechende strategische Wahl festgelegt worden waren, wurde das Projekt unter der Leitung eines Ausschusses, der vom Direktor der Hochschulbildung geleitet wurde, gestartet, wobei drei Phasen unterschieden werden können:

- eine Phase der Studie und der Analyse der Hauptmodelle des Verlagswesens (eine Studie, die dem *Institut des sciences du document numérique*, auf der Webseite des ENSSIB zugänglich, anvertraut wurde)<sup>19</sup>,
- eine Abstimmungsphase mit bevorzugten Partnern: Arbeitssitzungen mit den potentiell am Projekt Beteiligten (Redaktionen der Zeitschriften, Herausgeber); ausführliche Analyse von ERUDIT,
- eine Vorbereitungsphase des Aktionsplans, die mit einem Aufruf an die französischen Hochschulen zu einer Zusammenarbeit im Bereiche der Digitalisierung und der Verbreitung der Zeitschriftensammlungen nach deutlich aufgestellten Anforderungen abgeschlossen wurde.

Am Ende dieser Studien- und Vorbereitungsphase wurde am 7. März 2003 der Aufruf zur Zusammenarbeit veröffentlicht. Die Wahl der Einrichtung, die für das Endprojekt sorgen sollte, ist am 6. Juni 2003 erfolgt. Mit der Möglichkeit, allein oder im Konsortium zu antworten, haben mehrere öffentliche Einrichtungen beschlossen, gemeinsam einen Vorschlag auszuarbeiten. Insgesamt gingen vier Vorschläge ein. Nach Beratung des Auswahl Ausschusses wurde die Bewerbung der Universität von Lyon 2 in Partnerschaft mit der *Maison de l'Orient et de la Méditerranée* und der Universität von Nizza-Antipolis ausgewählt. Danach wurde ein Übereinkommen zwischen dem Staat, der Einrichtung, die für das Projekt sorgte, den Rechteinhabern der Zeitschriften und dem CINES<sup>20</sup> getroffen, um das Portal innerhalb eines Jahres nach der Unterzeichnung dieses Übereinkommens, die am 15. Oktober 2003 erfolgt ist, zu verwirklichen.

19 ENSSIB: Ecole nationale supérieure des sciences de l'information et des bibliothèques, <www.enssib.fr>. Institut des sciences du document numérique, vgl. <http://revues.enssib.fr>.

20 Das CINES, Centre informatique national de l'enseignement supérieur, ist Host der Webseite und für Langzeitarchivierung der Dateien verantwortlich. Das CINES liegt in Montpellier.

### Das Portal PERSEE

Infolge der Unterzeichnung des Übereinkommens konnte das Konsortium unter Leitung der Universität von Lyon 2 mit der Phase der Realisierung des Projekts beginnen. Das Projekt-Team wurde damit beauftragt, eine Digitalisierungsplattform aufzubauen und die Software des Portals zu entwickeln sowie die Verwaltung dieses Portals zu übernehmen.

Für diesen Auftrag sollten folgende Elemente geschaffen werden:

- Die sieben Zeitschriften sollten als Bild sowie als Text digitalisiert werden, vom ersten bis zum letzten Faszikel des Jahres 2002 (circa 410.000 Seiten);
- die Dateien sollten automatisch verarbeitet werden, um genormte XML-Notizen zu schaffen;
- die digitalisierten Dokumente sollten auf einer dafür spezifisch entwickelten Webseite online verfügbar gemacht werden;
- die Langzeitarchivierung der Dateien sollte vom CINES entwickelt werden.

Dem Projekt PERSEE ist es gelungen, innerhalb eines Jahres auf XML-Basis die erforderlichen technischen Mittel als Open Source Software – eine strategische Entscheidung, die durch das Ministerium getroffen wurde – zu entwickeln. Die technischen und dokumentarischen Lösungen machen die Besonderheit dieses Projekts aus. Sie erlauben, die zwei Recherchemöglichkeiten des Portals zu optimieren: *Browsen* und *Suchen*.

- Mit dem *Browsen* können die Benutzer/innen in derselben Zeitschrift von einem Faszikel zum anderen übergehen, die interaktiven Inhaltsverzeichnisse durchgehen und so den gewünschten Artikel finden. Die Digitalisierung der Zeitschrift als Bild erlaubt, die visuelle Identität der Zeitschrift zu respektieren, da jede befragte Seite mit dem Original auf Papier identisch ist.
- Bei der *Suche* liegt der Artikel im Zentrum der Recherche. Jeder Artikel wird neben einer Digitalisierung als Bild auch dank einem OCR als Text verfügbar gemacht und in XML nach einem Modell strukturiert, das erlaubt, die die Recherche betreffenden Daten (Name der Zeitschrift, Titel des Artikels, Name des Autors/der Autorin, Veröffentlichungsdatum, Zusammenfassung...) zur Geltung zu bringen.

Die Benutzer/innen können so eine präzise Suche in einem (oder mehreren) Gebiet(en) in einem bestimmten Feld machen oder diese auf den Volltext über die Gesamtheit des verfügbaren Corpus erweitern. Sie können auch andere Plattformen, mit denen Partnerschaften aufgestellt worden sind, befragen.

Die Wahl der Strukturierung und der Digitalisierung wurde von der Gemeinschaft der Forscher/innen getroffen, die an der Ausarbeitung des Lastenheftes des Projekts beteiligt waren. Die Verbindung eines ausgedehnten Corpus und einer mächtigen Suchmaschine begünstigt so die Transdisziplinarität, die den Geistes- und Sozialwissenschaften inhärent ist. Die zahlreichen Online-Dienste, die das PERSEE-Portal anbietet, wurden für die entsprechenden Bedürfnisse der Wissenschaftler/innen entwickelt. Außer den Browser- und Suchfunktionen werden auch personalisierte Dienste angeboten, die gemäß dem Profil des Nutzers/der Nutzerin aufgestellt wurden: persönlicher Raum, der eine Archivierung ausgewählter Artikel ermöglicht; automatische Bibliografieerstellung; Verwaltung des geschichtlichen Überblicks der Forschungsergebnisse; sofortige Information, wenn ein neuer interessanter Artikel online gestellt wird, usw. Die aktive Teilnahme der Redaktionen der Zeitschriften an diesem Prozess, sowohl in der Entscheidungsphase durch ihre Präsenz im Leitungsausschuss als auch in der konkreten Zusammenarbeit, was die materielle und intellektuelle Beschreibung ihrer Zeitschrift betrifft (um die Digitalisierungsoperationen und die Angemessenheit der Indexierung zu optimieren) und in der systematischen Suche nach den Autoren/innen oder ihren Rechteinhabern, muss auch erwähnt werden. Das französische Gesetz verpflichtet PERSEE nämlich, die Autoren/innen um eine offizielle Genehmigung zu bitten, um deren Werke online verfügbar machen zu dürfen. Die systematische Suche nach den Autoren/innen oder ihren Rechteinhabern, die von den Redaktionen der Zeitschriften verlangt wird, bildet also einen wesentlichen Teil des Programms PERSEE. In diesem Sinne konnte PERSEE in seiner Verwirklichungsphase von einer ständigen Rechtsbegleitung profitieren, die durch das CECOJI (*Centre d'Etudes sur la Coopération Juridique Internationale*) vom CNRS gewährleistet wurde, ein Zentrum, das auf geistiges Eigentumsrecht sowie auf das Recht der Infor-

mation und der Kommunikation spezialisiert ist.

#### *Die Entwicklungsperspektiven von PERSEE*

Die offizielle Eröffnung von PERSEE hat am 21. Januar 2005 das Ende der ersten Phase des Projekts markiert, die der Verwirklichung der Prototypen und dem progressiven Onlinestellen der Sammlungen der Pionierzeitschriften gewidmet war. Sie hat eine zweite Phase (2005-2007) eröffnet, deren Ziel die Entwicklung neuer Funktionen, die Einführung neuer Partnerschaften und die Integration neuer Zeitschriften ist. Parallel dazu wurde Ende 2005 damit begonnen, an eine Verdoppelung der Produktionskapazitäten zu denken, die dank der Öffnung eines zweiten Digitalisierungszentrums die Vermehrung der Zeitschriften auf PERSEE ermöglichen soll, da die potentielle Zahl der französischen Zeitschriften in den Geistes- und Sozialwissenschaften, und dadurch die potentielle Zahl der Zeitschriften auf PERSEE, sich auf über 200 beläuft.

Seit der Öffnung des Portals im Januar 2005 hat PERSEE ständig Inhalte hinzugefügt. Momentan verbreitet PERSEE die digitalisierten Sammlungen von zwölf Online-Zeitschriften mit einer Gesamtzahl von circa 30.000 Volltextartikeln und -berichten, alle kostenlos zugänglich, und ist damit die größte französischsprachige Webseite für die Verbreitung von Volltextartikeln in den Geistes- und Sozialwissenschaften. Während des Jahres 2005 wurden die Kandidaturen von 16 weiteren Zeitschriften von dem Leitungsausschuss akzeptiert, von fünf weiteren im März 2006, was die Gesamtzahl der Partnerzeitschriften auf 28 erhöht. Die Vorbereitung dieser neuen Zeitschriften ist jetzt im Gange, und die Online-Bereitstellung wird in den nächsten Monaten stattfinden. Für das Jahr 2006 sind noch sieben bis zehn weitere Titel geplant, so dass im Jahre 2007 mehr als 40 Zeitschriften online verfügbar sind.

PERSEE hat bis jetzt großen Erfolg beim Publikum erlebt. In knapp über einem Jahr wurden mehr als 20 Millionen Seiten heruntergeladen, circa 5.000 individuelle Benutzerkonten sind entstanden, der Durchschnitt liegt bei 200.000 Sitzungen pro Monat, der Höhepunkt im November 2005

lag bei mehr als 400.000 Verbindungen.<sup>21</sup> Die Tatsache, dass PERSEE für die Benutzer/innen vollständig kostenfrei ist und vielseitige Dienste anbietet, und dass PERSEE auch gut von den nationalen und internationalen Suchmaschinen und Portalen (SUDOC, *Google Scholar*, OAIster...) nachgewiesen wird, kann diesen Erfolg sehr wahrscheinlich erklären, genauso wie die einfache, benutzerfreundliche und effiziente Schnittstelle. Die Partnerschaften mit anderen Portalen von Zeitschriften in den Geistes- und Sozialwissenschaften werden die Attraktivität von PERSEE wahrscheinlich weiter erhöhen. Seit 2003 hat sich nämlich die Landschaft der französischsprachigen Portale von Zeitschriften in den Geistes- und Sozialwissenschaften stark entwickelt. Zu „den schon alten“ Projekten, wie ERUDIT in Quebec und Revues.org<sup>22</sup> in Frankreich sind öffentliche sowie private Initiativen gekommen.

Zu den privaten Initiativen gehört CAIRN<sup>23</sup>, ein Portal, das aus der Zusammenarbeit von vier belgischen und französischen Verlagshäusern stammt, die auf Geistes- und Sozialwissenschaften spezialisiert sind (Belin, de Boeck, La Découverte, Erès). CAIRN entstand im September 2005 und bietet zurzeit etwa 60 Zeitschriften als kostenpflichtiges Abonnement an. Eine weitere bemerkenswerte Initiative ist diejenige der Verlagshäuser Armand Colin und Sedes, die ein eigenes Portal für ihre laufende Produktion aufgebaut haben.<sup>24</sup> Was den öffentlichen Sektor angeht, muss man außer Revues.org und PERSEE das vom CNRS geführte CENS (*Centre d'édition numérique scientifique*) erwähnen. Das CENS, das im Dezember 2005 eröffnet wurde, ist noch in der Anfangsphase, was das Onlinestellen der Zeitschriften betrifft.<sup>25</sup>

Das französischsprachige Angebot im Bereich der Zeitschriften in den Geistes- und Sozialwissenschaften erscheint daher in mehrere Akteure aufgeteilt, die jeweils verschiedene wirtschaftliche Modelle verfolgen.

21 Seit der Redaktion dieses Beitrags in Februar 2006 ist die Zahl der Sitzungen auf bis zu 600.000 monatlich gestiegen (September 2006).

22 Vgl. <<http://www.revues.org>>.

23 Vgl. <<http://www.cairn.info>>.

24 Vgl. <<http://www.armand-colin.com/revues.php>>.

25 Vgl. <<http://www.cens.cnrs.fr/>>.

Außerdem drängt keine von diesen Portalen auf eine exklusive Verbreitung, das heißt, dass die Redaktionen der Zeitschriften die Wahl haben, auf mehr als einem einzigen Portal anwesend und zugänglich zu sein, insbesondere für ihre laufende Produktion, weil PERSEE zur Zeit das einzige Portal ist, das sich auf die retrospektive Sammlungen spezialisiert hat. Trotz dieser gesplitterten Lage haben sich diese Akteure in einem informellen Netz zusammengetan, in dem sie das Ziel verfolgen, über die technologische Interoperabilität ihrer Mittel und über die Verknüpfung zwischen ihren verschiedenen Angeboten nachzudenken, um einerseits die Kontinuität der auf den verschiedenen Portalen verbreiteten Sammlungen zu gewährleisten, andererseits es den Besuchern/innen zu ermöglichen, durch eine einzige Suche die Gesamtheit der verfügbaren Sammlungen auf diesen Portalen zu durchsuchen. Somit sollen die technologischen und dokumentarischen Lösungen, die vom Ministerium für PERSEE gefunden wurden, eine Verknüpfung zwischen den retrospektiven Sammlungen auf PERSEE und der laufenden Produktion der gleichen Zeitschrift auf einem anderen Portal erlauben.

Dem Konzept des freien Zugangs zu den retrospektiven Sammlungen der Zeitschriften – als Gegenleistung für die aus öffentlichen Mitteln finanzierte Digitalisierung – folgend, hat das Programm PERSEE eine wirkliche Umstrukturierung des Angebotes von Zeitschriften in den Geistes- und Sozialwissenschaften bewirkt; so konnte sich ein wirtschaftliches Modell etablieren, das auf einem Gleichgewicht zwischen einem freien und kostenlosen Zugang zu den retrospektiven Sammlungen (mit den von jeder Zeitschrift beschlossenen *Moving Walls*) und einem filterten und kostenpflichtigen Zugang zu den laufenden Veröffentlichungen der Zeitschriften beruht.

Dieses Portal, das den retrospektiven Sammlungen von Zeitschriften in den Geistes- und Sozialwissenschaften gewidmet ist, erweist sich als eine konkrete Antwort auf die Erwartungen der wissenschaftlichen Gemeinschaft. Als Frucht einer engen Kooperation mit den Forschern/innen, den Herausbergremien und den Fachleuten der Informationswelt, berücksichtigt PERSEE die derzeitigen Nutzungsmodelle, aber auch neue Bedürfnisse, um es der französischen Forschungsgemeinschaft zu erlau-

ben, von einer Verbreitung ihres wissenschaftlichen Kulturgutes angemessen zu profitieren. Von der *Direction de l'enseignement supérieur* und der *Direction de la Technologie* unterstützt, erlaubt PERSEE die Verbreitung eines außergewöhnlichen wissenschaftlichen Corpus im Netz, in einer Logik der Aufbewahrung, die den Wünschen der internationalen wissenschaftlichen Gemeinschaft entspricht.

### Schlussfolgerung

In mehreren Aspekten spiegelt sich die gesplitterte Landschaft der Digitalisierung in den Hochschulbibliotheken wider, einem Bereich, in dem Frankreich anderen Ländern wie Deutschland oder den USA gegenüber noch im Rückstand ist. Lokale Initiativen nehmen zu, manchmal mit einer Logik der reinen Aufbewahrung, manchmal in Verbindung mit den Themen der „Open Archives“ und der elektronischen Dissertationen. In diesem Zusammenhang ist es die Rolle des Ministeriums, die Projekte zu begleiten, die ihm unter finanziellem wie unter wissenschaftlichem Gesichtspunkt vorgelegt werden, und diese verschiedenen Initiativen zu koordinieren, um ihnen die bestmögliche Sichtbarkeit zu gewährleisten. Das SUDOC-Portal, das in der *Agence bibliographique de l'enseignement supérieur* (ABES) untergebracht ist und durch das Ministerium finanziert wird, trägt dazu bei, diese Sichtbarkeit zu erhöhen: dieses Portal soll also das Inventar der digitalisierten Sammlungen der Hochschulen aufnehmen.

In einer internationalen Lage, die von den Projekten großer privater Akteure (*Google, Yahoo, Microsoft*) oder öffentlicher Einrichtungen (der europäischen Kommission – Programm i2010) geprägt wird, entwickeln die französischen Hochschulbibliotheken ein digitales Angebot, das auf einen unvergleichbaren kulturellen und wissenschaftlichen Korpus zurückgreifen kann. Alle diese Initiativen werden dazu beitragen, langfristig die linguistische und kulturelle Vielfalt der wissenschaftlichen Dokumentation im Internet aufrechtzuerhalten und zu erweitern.

\*\*\*

*Frédéric Blin ist im Ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche in der Direction générale de l'enseignement supérieur, Sous-direction des bibliothèques et de l'information scientifique, beschäftigt. Er ist für die Internationale Kooperation zuständig. E-Mail: frederic.blin@education.gouv.fr*

*Valérie Néouze ist ebenfalls dort beschäftigt und für den Bereich Digitalisierung zuständig. E-Mail: valerie.neouze@education.gouv.fr*

## KOPAL – EIN DIGITALES ARCHIV ZUR DAUERHAFTEN ERHALTUNG UNSERER KULTURELLEN ÜBERLIEFERUNG

von Thomas Wollschläger

*Der Wandel des Publikationsmarktes hin zur elektronischen Publikation in Virtuellen Netzen verändert das Aufgabenprofil insbesondere der Bibliotheken, die für die Bewahrung, Sicherung und Überlieferung des kulturellen Erbes einer Nation mitverantwortlich sind, erheblich. Ziel und Zweck des durch das Bundesministerium für Bildung und Forschung (BMBF) geförderten Projekts kopal<sup>1</sup> ist die praktische Erprobung und Implementierung eines kooperativ erstellten und betriebenen Langzeitarchivierungssystems für digitale Dokumente und Daten als eine nachnutzbare Lösung für sowohl die Langzeiterhaltung als auch die Sicherstellung der Langzeitverfügbarkeit digitaler Ressourcen. Der Artikel dokumentiert die Schwerpunkte und den Projektverlauf des Vorhabens kopal zwischen 2004 und 2007.*

\*\*\*

### Herausforderungen der digitalen Langzeitarchivierung

Der internationale und auch der deutsche Publikationsmarkt ist geprägt von einem weiterhin rasch wachsenden elektronischen Publikationssektor. Sowohl das wissenschaftliche Publizieren als auch allgemein gesellschaftliche und kulturelle Aktivitäten finden mit einem zunehmenden Anteil im Netz statt. Heute angewandte und machbare Selektionskriterien bei der Sammlung elektronischen Materials entscheiden über die Tradierung unseres kulturellen Erbes. Dieser „Nachlass“ ist – auch gesetzlich fixiert – eine Aufgabe von National- und Landesbibliotheken und vergleichbaren

<sup>1</sup> „Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen“; Projekt-Website <<http://kopal.langzeitarchivierung.de/>> (27.02.2006).

Einrichtungen. Auch die aktuelle Novellierung des Gesetzes über Die Deutsche Bibliothek steht in diesem Kontext: Sie erweitert den Sammelauftrag der künftigen Nationalbibliothek<sup>2</sup> um nicht-trägergebundene digitale Materialien. Teile dieser Materialien, wie Online-Hochschulschriften, Online-Newsletter oder elektronische Zeitschriften einiger Verlage, sammelt Die Deutsche Bibliothek seit Jahren auf freiwilliger Basis. Der neue Auftrag bezieht nunmehr sowohl sämtliche Formen elektronischer Publikationen ein als auch einen enorm gewachsenen Umfang dieser Materialien (zum Beispiel im Rahmen von Web-Archiving).

Um dem erweiterten Sammelauftrag gerecht zu werden, benötigen die Bibliotheken und Gedächtnisorganisationen entsprechend geeignete und dimensionierte Archive für elektronische Materialien. Die bestehenden Archive erfüllen jedoch die Anforderungen an „vertrauenswürdige Archive“<sup>3</sup> nicht oder nur teilweise. Im Fokus der Bemühungen, ein geeignetes Archivsystem für das elektronische Material zu finden, standen unter anderem folgende Überlegungen:

- Als Grundlage müssen die binären Daten erhalten werden, denn kein existierender Datenträger ist ewig oder auch nur langfristig genug haltbar. Das Archiv müsste also regelmäßige Erhaltungsmaßnahmen (Datenträgermigrationen) durchführen können.
- Der rasante Technologiewechsel erschwert immer wieder aufs Neue den Zugriff auf ältere Datenformate. So gibt es bereits zahlreiche Formate, zu denen ständig neue hinzukommen, während bisherige Formate verschwinden oder obsolet werden. Somit entstehen komplexe Abhängigkeiten von aktueller Soft- und Hardware. Das Archiv müsste daher in der Lage sein, sowohl Format-Migrationen (regelmäßiges Konvertieren) als auch Emulationen (Nachstellen benötigter Systeme) zu ermöglichen.

Bei der Migration werden alte Dateiformate rechtzeitig, das heißt solange

<sup>2</sup> Die Gesetzesnovellierung ist mittlerweile zum 29.06.2006 in Kraft getreten. Die Deutsche Bibliothek heißt nunmehr Deutsche Nationalbibliothek (DNB).

<sup>3</sup> „Trusted Digital Repositories“. Siehe hierzu vor allem die Materialien der nestor-AG „Vertrauenswürdige Archive“, <<http://nestor.cms.hu-berlin.de/tiki/tiki-index.php?page=AG%20Vertrauensw%FCrdige%20Archive>> (27.02.2006).

man die Formate noch vollkommen lesen kann, in aktuellere Formate konvertiert. Dies erfolgt periodisch und bedeutet eine Überführung aus bestimmten Hardware- und Software-Konfigurationen in eine neuere Umgebung, die aktuell technisch und organisatorisch beherrscht werden kann. Zweck der Migration ist es also, die Integrität und die Verfügbarkeit digitaler Ressourcen trotz des stetigen Wandels der technischen Umgebung zu erhalten. Problematisch bei diesem Verfahren ist jedoch der möglicherweise nicht erkannte Verlust von (Teil-)Informationen bei automatisch ablaufenden Migrationsroutinen; je komplexer das vorliegende Datenformat ist, desto größer wird die Wahrscheinlichkeit, dass bei längeren Migrationsketten (über Jahre hinweg) bestimmte Features nicht mehr vorhanden bzw. lauffähig sind. Demgegenüber ist die Migration ein geeignetes Verfahren für große Datenmengen, und es dürfte vor allem für eher statische Formate die vernünftigste Strategie sein.

Bei der Emulation versucht man mit einem speziellen Programm (dem Emulator) oder gegebenenfalls mit mehreren Programmen auf einer aktuellen Systemumgebung eine alte Systemumgebung nachzubilden. Zweck der Emulation ist es also, auf einem (aktuellen) System Daten und Programme zu verarbeiten, die ursprünglich für ein anderes, historisches System bestimmt waren. Dies kann sehr aufwändig sein und es setzt vor allem eine sehr genaue Definition der hard- und softwareseitigen Systemanforderungen voraus. Ein Vorteil der Emulation ist es jedoch, dass sie für sehr komplexe Formate (zum Beispiel Multimedia-Anwendungen) eine geeignete Lösung darstellt, um die Features dieser Formate auf lange Zeit nutzbar zu halten.

Beide Herangehensweisen haben also Vor- und Nachteile, weshalb ein Archivsystem in der Praxis zur Langzeiterhaltung und Langzeit-Verfügbarmachung digitaler Publikationen eine Kombination beider Verfahren einsetzen können müsste.

### Grundprinzipien des kopal-Projekts

Das Projektziel von kopal ist daher der Aufbau einer von Gedächtnisorganisationen nachnutzbaren technischen und organisatorischen Infrastruktur zur Sicherung der Langzeitverfügbarkeit elektronischer Publikationen. Es

ist ein Förderprojekt des Bundesministeriums für Bildung und Forschung (BMBF) über eine Laufzeit von drei Jahren (bis Mitte 2007, Fördervolumen: 4,2 Millionen Euro). Den Kern des kopal-Archivsystems bildet das von IBM in Zusammenarbeit mit der Nationalbibliothek der Niederlande (Koninklijke Bibliotheek, KB) entwickelte Digital Information Archiving System (DIAS)<sup>4</sup>. Innerhalb des Projekts werden digitale Materialien aller Art der DDB und der Niedersächsischen Staats- und Universitätsbibliothek (SUB Göttingen) im Massenverfahren in das Langzeitarchiv eingestellt.

Im Wesentlichen beruht kopal dabei auf den drei Prinzipien *Kooperation*, *Universalität* und *Nachnutzbarkeit*. Im Rahmen einer nationalen Kooperation arbeiten im Projekt zwei sehr unterschiedliche Bibliotheken zusammen. Dies betrifft nicht nur den jeweiligen Auftrag (Nationalbibliothek bzw. Sondersammelgebietsbibliothek der DFG), sondern auch Schwerpunkte bei den bisherigen Aktivitäten (Online-Dissertationen und Netzpublikationen bei der DDB, Digitalisate und im naturwissenschaftlich-technischen Bereich gängige Datenformate bei der SUB Göttingen). Auf diese Weise werden sehr unterschiedliche Aspekte in das Projekt eingebracht, die die Erfahrungspalette spürbar erweitern und die Nachnutzungsmöglichkeiten für weitere Nutzer/innen nochmals deutlich erhöhen. Außerdem arbeiten die Projektpartner eng mit dem „Kompetenznetzwerk Langzeitarchivierung“ (nestor)<sup>5</sup> zusammen. Der technische Betrieb des Langzeitarchivs ist beim Rechenzentrumspartner „Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen“ (GWDG) angesiedelt. Der Entwicklungspartner IBM Deutschland GmbH ermöglicht eine professionelle Anpassung der Softwarekomponenten und bietet eine langfristig stabile Unterstützung.

Im internationalen Rahmen arbeiten die Projektpartner zum einen eng mit der Königlichen Bibliothek der Niederlande zusammen, die das originale DIAS-Archivsystem als In-House-Lösung betreibt. Zusammen mit ihr werden Anforderungen an künftige Weiterentwicklungen von DIAS sowie

4 Siehe <<http://www-5.ibm.com/nl/dias/index.html>> (27.02.2006).

5 Ausführliche Informationen unter: <<http://www.langzeitarchivierung.de/>> (27.02.2006).

Strategien eines „Preservation Planning“ entwickelt. Außerdem beteiligt sich kopal an der internationalen Entwicklung und dem Austausch von Metadaten zur Langzeitarchivierung und kooperiert mit verschiedenen internationalen Projekten zur Langzeitarchivierung.

Das Projekt erfüllt den Anspruch an ein universell nutzbares Archivsystem, indem zum einen die Sicherstellung der langfristigen Verfügbarkeit durch Migration und Emulation unterstützt wird. Zum anderen gibt es dabei in kopal keinerlei Einschränkungen sowohl für die Art des Materials, welches in das Archiv eingespielt werden kann (Text, Bilder, Audio, Video), als auch für die Bandbreite der möglichen Dateiformate (PDF, HTML, TIFF usw.). Wiewohl das kopal-System für den Projektzeitraum eine begrenzte Gesamtkapazität hat, ist die Größe des einzelnen Archivobjekts nicht begrenzt. Jeder der Partner ist zunächst völlig frei in der Auswahl und Regelfestlegung beim Einspielen der von ihm gesammelten Objekte. Gleichzeitig ist es möglich, unter Einbeziehung des betreibenden Partners GWDG Regeln für die gezielte Migration einzelner Objektgruppen zu definieren und sie auch über alle Bestände im Archiv umzusetzen.

Ein wesentlicher Aspekt des Systems ist schließlich die dezidierte Nachnutzbarkeit. Um diese zu gewährleisten, werden etablierte Standards genutzt. Der Transfer der Objekte in ein digitales Archiv über standardisierte Formate, Transportwege und Systemschnittstellen ist dabei ein wichtiges Erfordernis. Ohne Standardisierung ist die Bewältigung großer Mengen von Objekten nicht möglich und ein arbeitsteiliges Zusammenwirken kooperierender Archivsysteme wäre deutlich erschwert.

Das Projekt kopal hat dafür mit dem „Universellen Objektformat“ ein Austausch- und Archivformat vorgelegt und implementiert, mit dem digitale Objekte zusammen mit Metadaten archiviert und zwischen Institutionen und Archivsystemen ausgetauscht werden können. Es basiert auf den Formaten METS<sup>6</sup> in der Version 1.4 und den Langzeitarchivierungsmetadaten für elektronische Ressourcen (LMER)<sup>7</sup> in der Version 1.2. Ein

6 Metadata Encoding & Transmission Standard; <<http://www.loc.gov/standards/mets/>> (27.02.2006).

7 Informationen unter: <<http://www.ddb.de/standards/lmer/lmer.htm>> (27.02.2006).

Beispiel für die Metadatenstruktur eines Archivobjekts in kopal gibt die folgende Abbildung wieder:

```
<?xml version="1.0" encoding="utf-8" standalone="no" ?>
< mets OBJID="1" TYPE="Kopal Submission Information Package" PROFILE="DDB" xmlns="http://www.loc.gov/METS/" xmlns:slink="http://www.w3.org/1999/xlink"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:lmer="http://www.ddb.de/LMERfile" xmlns:lmerObject="http://www.ddb.de/LMERObject"
xmlns:lmerProcess="http://www.ddb.de/LMERprocess" xsi:schemaLocation="http://www.loc.gov/METS/ http://www.loc.gov/standards/mets/mets.xsd
http://www.ddb.de/LMERfile http://www.ddb.de/standards/lmer/lmer-file.xsd http://www.ddb.de/LMERObject http://www.ddb.de/standards/lmer/lmer-
object.xsd http://www.ddb.de/LMERprocess http://www.ddb.de/standards/lmer/lmer-process.xsd" />
<!-- METS CREATE DATE = 2006-02-16T10:26:25 RECORD STATUS = PRODUCTION -->
<!-- AGENT ROLE = ARCHIVIST TYPE = ORGANIZATION -->
<!-- NAME = Deutsche Bibliothek -->
<!-- NOTE = Automatisch generierte Metadaten. Für weitere Informationen: kopal.langzeitarchivierung.de -->
</agent>
<!-- METS --÷ data-bbox="578 622 954 709" data-label="Text">


Abbildung 1: Beispiel für eine mets.xml-Datei, die Metadaten zum Archivobjekt im Universellen Objektformat abbildet. Es handelt sich in dem Fall um eine Online-Dissertation, bestehend aus den Dateien Hauptdokument, Abstract und Checksumme.


```

Zum anderen wird die zuverlässige Kern-Software durch flexible Open Source-Module erweitert. Das kopal-Archivsystem besteht aus dem weiterentwickelten DIAS der IBM und den von den Partnern SUB Göttingen und der DDB entwickelten kopal-Tools. Dabei beruht DIAS auf Standardsoftware-Komponenten der IBM und richtet sich in seiner Architektur

streng nach dem OAIS-Modell.<sup>8</sup> Hierfür erstellen die DDB und die SUB Göttingen auf DIAS abgestimmte Softwareprodukte: die kopal-Tools, die als „kopal Library for Retrieval and Ingest“ (koLibRI) unter einer Open Source-Lizenz veröffentlicht werden. Die Systementwicklung wird dabei so offen angelegt, dass eine Ausdehnung der kooperativen Nutzung um weitere Archivbibliotheken sowie Nachnutzer aus dem Kreis aller „Gedächtnisorganisationen“ (also auch Archive und Museen) und wissenschaftlichen Datenarchive möglich ist.

Die Erweiterbarkeit von DIAS für neue Nutzerinstitutionen und präzise definierte Import- und Exportschnittstellen ermöglichen es, eine Archivnutzung in verschiedenste Workflows unterschiedlichster Institutionen zu integrieren. Durch den Einsatz von vielfach bewährter IBM-Standardsoftware, wie der DB2-Datenbank, dem Content Manager und dem Tivoli Storage-Manager, sind langfristige Stabilität, Performanz und Skalierbarkeit gewährleistet.

Die Deutsche Bibliothek und die Staats- und Universitätsbibliothek Göttingen entwickeln zurzeit ein Software-Paket für die Benutzung der kopal-Solution: die „kopal Library for Retrieval and Ingest“ (koLibRI). Bei diesen *kopal-Tools* geht es hauptsächlich um den Bereich des Einspiels von Objekten in das DIAS sowie um den Zugriff auf die archivierten Objekte (siehe Abbildung 2).

<sup>8</sup> OAIS = Open Archival Information System (ISO Standard 14721). DIAS verwaltet – dem OAIS-Standard für Archivsysteme entsprechend – die gespeicherten Daten und stellt eine Umgebung für unterschiedlichste Erhaltungsstrategien zur Verfügung. Dabei wird das Objekt als Submission Information Package (SIP) verarbeitet und als Dissemination Information Package (DIP) geliefert (Terminologie gemäß dem OAIS-Modell). Innerhalb des DIAS wird ein Datenteil (Archival Information Package oder auch AIP) abgetrennt und gelangt in einen Storage (Festplatte, Magnetband usw.). Bestimmte Metadaten werden in einer Datenbank (Data Management) abgelegt, auf die über eine Administrationschnittstelle zugegriffen werden kann. Zum OAIS-Standard siehe <<http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>> (27.02.2006).

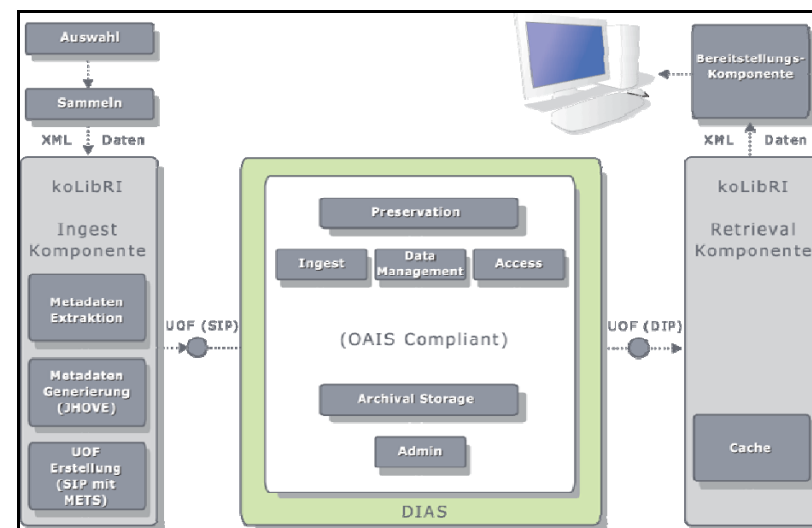


Abbildung 2: Der Aufbau des kopal-Archivsystems richtet sich nach dem OAIS-Modell. Die koLibRI-Software übernimmt die Erstellung der Archivobjekte und deren späteres Retrieval.

Die Schnittstelle für den Datenimport (Ingest) erfüllt folgende Anforderungen:

- Der Ingest kann automatisiert erfolgen.
- Das universelle Objektformat (kopal-UOF) wird unterstützt.
- Die flexible Schnittstelle ermöglicht eine Integration in verschiedenste Umgebungen und Informationssysteme.
- Es werden internationale Standards angewandt.
- Eine Nachnutzbarkeit durch Dritte ist gewährleistet.
- Eine grafische Oberfläche ist in Entwicklung.

Die Anforderungen an den Datenexport (Access) sind bei der DDB und der SUB Göttingen sehr unterschiedlich, so dass generische Module genutzt werden, die von den einzelnen Institutionen je nach Bedarf erweitert werden können. Mit Hilfe dieser Software soll künftig auch eine Administration des kopal-Systems realisiert werden.

Aufgrund der sehr unterschiedlichen und teils heterogenen Systemstrukturen der beiden Institutionen wird eine flexible Software benötigt. Das so genannte Workflow-Tool bietet eine gemeinsam nutzbare Infrastruktur für Module. Dieses Workflow-Tool kann als Asset-Builder zum Erstellen von Archivpaketen dienen. Weiterhin kann es als eine zentrale Vermittlungsinstanz zum DIAS-System, als Client-Loader, genutzt werden, indem es Archivpakete von mehreren Asset-Buildern sammelt und an DIAS übermittelt. Weitere Nutzungsmöglichkeiten lassen sich zudem problemlos integrieren.

### **Der bisherige Projektverlauf von kopal**

Nach einer Pilotphase zur Systemevaluierung und Entwicklung des Universellen Objektformats befindet sich kopal derzeit inmitten der Entwicklungsphase, die auf verschiedene Arbeitspakete aufgeteilt worden ist. Die DIAS-Basisentwicklung und -Bereitstellung sowie die DIAS-Schnittstellen-Implementierung sind dabei bereits abgeschlossen worden. Die Kernkomponente DIAS wurde durch IBM an die Anforderungen der Partner angepasst. Verantwortlich für die Bereitstellung des erforderlichen Hardwareumfeldes in dieser Phase war die GWDG. Die Einrichtung der Mandantenfähigkeit hat zusätzlich einen flexiblen und skalierbaren Zugriff auf das System gesichert und der Fernzugriff lässt nun auch einen nicht ortsgebundenen Zugang eines Mandanten auf das System zu. Dadurch wurde auch eine zukünftige Aufnahme neuer Teilnehmer/innen ermöglicht.

Die Partner DDB und die SUB Göttingen haben auf der Basis des DIAS Software implementiert, die die Voraussetzung für ein Überführen bestehender und zukünftiger Bestände sowie für Zugriffe auf das Langzeitarchiv geschaffen hat. Diese äußeren Dienste sind insbesondere zur Nachnutzung durch Dritte gedacht und deshalb an internationalen Standards ausgerichtet.

Ende 2005 erfolgte eine Testphase zur Abnahme des bei der GWDG installierten Systems und für die Erprobung der von der DDB und der SUB Göttingen entwickelten Ingest- und Retrieval-Tools. Beide Bibliotheken haben Testobjekte aus ihren elektronischen Beständen in die Tests eingebracht. Die GWDG stellte die Testumgebung zur Verfügung und führte

Tests bezüglich betriebsüblicher Störungen und Systemausfälle durch. Die Testphase verlief erfolgreich und das System wurde schließlich durch das Projekt kopal abgenommen.

In der aktuellen Projektphase (Stand: Februar 2006) ist nun auch das Aufsetzen des Produktivsystems abgeschlossen worden. In den kommenden Monaten spielen DDB und die SUB Göttingen eine Auswahl ihrer vorhandenen Datenbestände ein. Ziel ist es, anhand einer Vielzahl von Objekten in verschiedenen Formaten die praxisnahe Nutzung der kopal-Lösung aufzuzeigen. Gleichzeitig werden bei Projektende die exemplarisch übernommenen Datenbestände den Grundstock der dauerhaften Archivierung der elektronischen Materialien der Bibliotheken bilden.

### **Einzuspielende Daten für kopal**

Die beiden Bibliotheken in Frankfurt und Göttingen werden im Rahmen der exemplarischen Datenübernahme Bestände einspielen, die sowohl bisher auf freiwilliger Basis gesammelte Online-Publikationen umfassen als auch verschiedene digitalisierte Quellen, CD-Rom-Images und weitere Datenarten.

Für die DDB ist die Übernahme folgender Daten bzw. von Teilen folgender wissenschaftlich und kulturell bedeutsamer Datenbestände vorgesehen:

- Online-Dissertationen (Originär digitale Dokumente ab 1997); Anzahl: derzeit 41.000; Datenmenge: 300 Gigabyte
- Netzpublikationen (elektronische Zeitschriften, zum Beispiel SpringerLINK); Datenmenge: 300 Gigabyte
- CD-ROMs (Quellensammlungen, Multimedia-Anwendungen, Filme, Spiele, Datensammlungen (zum Teil wissenschaftliche Primärdaten), sonstige Publikationen); Anzahl: 25.000; Datenmenge: 14.300 Gigabyte
- Digitalisate der Herzog-August-Bibliothek Wolfenbüttel; Datenmenge: 1.500 Gigabyte
- Exilpresse Digital und Jüdische Zeitschriften in NS-Deutschland (Bestand der Deutschen Bibliothek, Frankfurt); Datenmenge: (vorerst) 150 Gigabyte

- Digitalisate des Deutschen Buch- und Schriftmuseums, Leipzig; Datenmenge: (vorerst) 10.000 Gigabyte
- Digitale Musikdaten (von Schellackplatten bis Notendigitalisaten, Bestand des Deutschen Musikarchivs, Berlin); Datenmenge: 544.000 Gigabyte

Bis auf die CD-Rom-Images und die Digitalen Musikdaten sollen diese Datenbestände im Projektzeitraum möglichst vollständig nach kopal übernommen werden. Dabei hat sich im bisherigen Projektverlauf gezeigt, dass die ursprünglich für kopal geplanten Kapazitäten trotz der damals sehr großzügigen Anlage des Systems selbst für die bis Ende 2007 vorhandenen Materialien nicht ausreichen werden. Einen Eindruck des Verhältnisses von Materialmengen, die eingespielt werden könnten, und den geplanten Kapazitäten gibt die Abbildung 3 wieder.

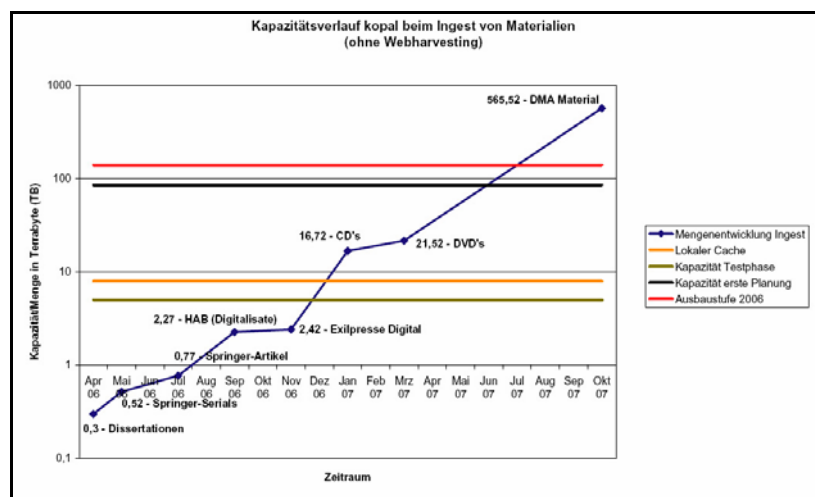


Abbildung 3: Geschätzter Mengenzuwachs bei einzuspielenden Materialien in kopal in Beziehung zu den bei Beginn des Projekts geplanten Kapazitätsgrenzen (logarithmische Darstellung)

Die Grenzen für den lokalen Cache spielen nur für die Optimierung der

Performanz des Systems beim Abruf von Daten eine Rolle und sind daher für das Einspielen der Daten kein Hindernis. Um die maximale Kapazität des bei der GWDG installierten Systems zu erhöhen und um so viele Materialien wie möglich übernehmen zu können, plant man derzeit, einen erheblichen Speicherzuwachs vorzunehmen. Dies ist umso wichtiger, da ein Bereich in der oben dargestellten Architektur noch nicht eingeschlossen ist, nämlich die Archivierung geharvesteter Websites.

Im Jahre 2005 hat die Australische Nationalbibliothek ein Domain-Komplett harvesting für die Domain .au (= Australien) durchgeführt. Dabei wurden rund 185 Millionen eindeutig unterscheidbare Dokument(seiten) von 811.000 Sites geharvestet, mit einem Gesamtvolumen von 6,69 Terabyte (in komprimierter Form zuzüglich Archivierungsmetadaten: 4,6 Terabyte).<sup>9</sup> Dies zeigt, dass selbst für das einmalige Harvesten einer kleineren nationalen Domain enorme Datenmengen anfallen. Auch die DDB hat zur Erprobung von Harvesting-Strategien bereits Pilotmaßnahmen durchgeführt. Es werden derzeit Strategien entwickelt, welche den Umfang der zu harvestenden Domain(s) bzw. Sites für Deutschland, die Frequenz des Harvesting und die Frage von gezieltem oder generellem Harvesting adressieren.

### Herausforderung: Preservation Planning und Service

Neben der exemplarischen Datenübernahme steht derzeit die Adressierung des Arbeitspakets Preservation Planning und Bitstream-Preservation. Die Partner IBM, DDB und SUB Göttingen entwickeln dabei unter Einbeziehung der KB (Niederlande) ein tragfähiges Konzept zum Preservation Planning. Dabei geht es um die Erkennung der digitalen Objekte, die bedingt durch den technologischen Wandel in der Zukunft nicht mehr zugänglich sein werden. Dazu gehören die Anbindung einer funktionierenden internationalen Format Registry, die performante Migration großer Datenmengen und die erfolgreiche Umsetzung von Emulationsmechanismen.

<sup>9</sup> Bericht unter: <[http://pandora.nla.gov.au/documents/domain\\_harvest\\_report\\_public.pdf](http://pandora.nla.gov.au/documents/domain_harvest_report_public.pdf)> (27.02.2006).

Das entwickelte Konzept soll innerhalb der kopal-Solution umgesetzt werden. Eine Grundlage dafür ist die Bitstream-Preservation, das sichere und verlustfreie langfristige *Erhalten* von digitalen Objekten. Die dafür nötigen Speicherstrategien sowie die Migrations- und Emulationsstrategien, welche das eigentliche Preservation Planning im Sinne der langfristigen *Verfügbarmachung* ausmachen, werden von den Projektpartnern entwickelt.

Um eine sichere Speicherung und langfristige Verfügbarkeit digitaler Daten zu gewährleisten, muss die Archivierungslösung in bestehende Informationssysteme und Arbeitsabläufe integriert werden. Hierbei ist die Nachnutzung durch weitere Kulturerbeinstitutionen sowie durch sonstige Institutionen, die eine Langzeitarchivierung benötigen, ausdrücklich erwünscht.

Die kopal-Solution ist von vornherein auf unterschiedliche Bedürfnisse ausgerichtet. Kopal verfolgt dabei das Ziel, künftig verschiedene Nutzungsmodalitäten bereitzustellen: Es wird einerseits die Möglichkeit geboten werden als Mandant mit einem eigenen „Schließfach“ das bestehende System mit abgesichertem Speicherplatz und eigener Verwaltung der Daten zu nutzen. Diese Lösung ist besonders geeignet für Institutionen mit geringerem Archivgutaufkommen. Andererseits wird die Möglichkeit einer Nachnutzung der kopal-Solution durch eine eigene Installation des DIAS-Systems bestehen. Dies wird in der Regel zusammen mit einer Nutzung der koLibRI-Software erfolgen.

### **Perspektive**

Bereits seit März 2006 steht für Testzwecke ein vorläufiger Release der neu erstellten Ingest- und Retrieval-Software (kopal-Tools bzw. koLibRI) in einer Beta-Version zur Verfügung. Auf der Website von kopal gibt es – neben weiteren Informationen zum Projekt, den Standards und Downloads von Dokumentationen – ebenfalls seit März 2006 den „kopal-Demonstrator“ – eine interaktive Animation, die die Grundlagen, Funktionalitäten und Module des Systems anschaulich darstellt.

Gemeinsam mit der Königlichen Bibliothek der Niederlande haben die kopal-Partner eine DIAS-User-Group ins Leben gerufen. Die beteiligten

Institutionen halten mehrmals im Jahr Arbeitstreffen ab, um Informationen und Erfahrungen auszutauschen, notwendige Anpassungen zu diskutieren und gemeinsame Strategien für künftige Nutzenanwendungen zu entwickeln. Außerdem steht dem Projekt ein Beirat mit Experten aus Wirtschaft, Wissenschaft und öffentlicher Verwaltung zur Seite, die den Gesamtfortschritt und Ablauf des Projektes beratend begleiten. Die Zusammensetzung des Beirats gewährleistet, dass die Ziele von kopal gut unterstützt und die strategische Zielsetzung erreicht werden kann.

Mit dem Projektabschluss im Juni 2007 werden ein nachnutzbares System und ein endgültiges Release der vollständig entwickelten koLibRI-Software bereitgestellt. Hierauf bauen die unterschiedlichen kopal-Dienste auf.

\*\*\*

*Dr. Thomas Wollschläger ist Wissenschaftlicher Mitarbeiter im Projekt kopal in der Abteilung Informationstechnik der Deutschen Bibliothek Frankfurt am Main. Innerhalb des Themenkomplexes Langzeitarchivierung elektronischer Daten und speziell des Projekts sind seine Aufgabenschwerpunkte die Vernetzung des Archivsystems mit den Workflows für elektronische Materialien innerhalb der Bibliothek, die Koordination der Fachabteilungen und externer Partner sowie die Unterstützung des Projektcontrollings. E-Mail: t.wollschlaeger@d-nb.de*

DAS PROJEKT NESTOR – KOOPERATIVE STRUKTUREN  
FÜR DIE ÜBERLIEFERUNGSBILDUNG – BEISPIELE AUS DEM  
BESTAND MASCHINENLESBARER DATEN DES BUNDESARCHIVS

von Andrea Hänger und Karsten Huth

*Der Aufsatz beschreibt den Beitrag des Bundesarchivs im Rahmen des Projekts nestor. Zunächst werden das Projekt und der Bestand maschinenlesbarer Daten des Bundesarchivs vorgestellt. Dann wird der aktuelle Stand der Forschung anhand von konkreten Beispielen dargestellt. Zur Sprache kommen Fragen zur Vertrauenswürdigkeit von Archiven, zur Authentizität von digitalen Objekten und zu Metadaten für die Langzeitarchivierung (PREMIS).*

\*\*\*

### Einleitung

In seinem Artikel „Ensuring the Longevity of Digital Documents“ beginnt der Autor Jeff Rothenberg mit einer kurzen Geschichte, die in der Zukunft spielt. Seine Enkel finden im Jahr 2045 einen Brief und eine CD ihres verstorbenen Großvaters. Der Brief besagt, dass auf der CD der Weg zum Erbe des Großvaters beschrieben ist. Leider können seine Enkel die CD nicht verwenden, weil es schon seit dreißig Jahren keine CD-Laufwerke mehr gibt und die entsprechende Software, die das Testament lesbar machen könnte, schon längst nicht mehr verfügbar ist. Somit ist das Erbe für immer verloren.<sup>1</sup> Dieser Artikel von 1995 stammt aus einer Zeit, in der der technologische Fortschritt scheinbar grenzenlos schien. Dennoch markiert er auch den Anfangspunkt, an dem sich weite Kreise von Fachleuten der

<sup>1</sup> Rothenberg, Jeff, Ensuring the Longevity of Digital Documents, in: Scientific American 272 (1995), S. 24-29, hier S. 24.

Möglichkeit eines digitalen Desasters für die zukünftige Überlieferungsbildung bewusst wurden. Es folgten zehn Jahre an kooperativen Bemühungen, um Lösungen und Strategien für das Problem der digitalen Langzeitarchivierung zu finden. Dieser Artikel soll an Beispielen aus der Praxis zeigen, wie Kooperationen neue Entwicklungen auf dem Gebiet der historischen Überlieferungsbildung fördern können.

Die Langzeitarchivierung digitaler Objekte ist eine Herausforderung, der sich nicht nur die klassischen Gedächtnisorganisationen wie Archive, Bibliotheken und Museen stellen müssen, sondern auch die Produzenten und Dienstleister. Viele der wissenschaftlichen, kulturellen und administrativen Ressourcen werden heute ausschließlich digital produziert. Hinzu kommen die Ressourcen aus Digitalisierungsprojekten, die mit erheblichem Aufwand, zum Schutz der analogen Objekte, ins Leben gerufen worden sind. Das Problem der digitalen Langzeitarchivierung ist deshalb so komplex, weil man es mit einer Fülle von unterschiedlichen Objekten zu tun hat. Von reinem Text hin zu digitalisierten, hochaufgelösten Bildern, hin zu Datenbanken und wissenschaftlichen Rohdaten, komplexen Webseiten, elektronischen Akten, E-Journals, Multimedia und Computerprogrammen. Hinzu kommen komplizierte rechtliche Bestimmungen des Urheberrechts. Aufgrund der Komplexität des Problems sind kooperative Strukturen bei der Lösungsfindung unabdingbar.

### Die kooperative Struktur – nestor

Eine solche kooperative Struktur stellt das Network of Expertise in Long-Term Storage of Digital Resources (nestor), das „Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit Digitaler Ressourcen“, dar.<sup>2</sup> „Nestor wurde initiiert,

- um diejenigen zusammenzubringen, die sich hierzulande mit dem Problem der dauerhaften Erhaltung digitaler Objekte beschäftigen,
- um eine Informations- und Kommunikationsplattform als zentrale Anlaufstelle für alle Fragen der digitalen Langzeitarchivierung zu schaffen und

<sup>2</sup> Vgl. <<http://www.langzeitarchivierung.de>>.

- um einen nach außen sichtbaren Focus in Deutschland zu bilden, der Kooperationen vermittelt und als Einstiegspunkt für internationale Allianzen zur Verfügung steht.
- Langfristiges Ziel von nestor ist der nachhaltige Aufbau einer kooperativen Infrastruktur, in der vielfältige Fachkompetenzen zusammenwirken.<sup>3</sup> Das Bundesarchiv ist seit 2005 ein Partner von nestor.

### Maschinenlesbare Bestände des Bundesarchivs

Das Bundesarchiv beschäftigt sich seit vielen Jahren mit dem Thema der Archivierung elektronischer Unterlagen. Die ältesten archivierten Daten stammen aus dem Jahr 1970. Sie sind heute noch vollständig lesbar und interpretierbar. 35 Jahre sind für Archive/innen eine ausgesprochen kurze Zeitspanne. Für die Zeitrechnung der Informationstechnologie hingegen ist diese Zeit schon beachtlich. Rund 200 digitale Archivobjekte mit ungefähr 300.000 Dateien, die im Bundesarchiv verwahrt werden, sind in der Deutschen Demokratischen Republik in der Zeit von 1970 bis 1990 entstanden. Datenbanken aus den Bereichen Statistik, Wirtschaft, Landwirtschaft, Bildung, Strafvollzug und Arbeit konnten erhalten werden. Diese Daten werden regelmäßig von wissenschaftlichen Benutzern/innen, aber auch von Betroffenen genutzt, die mit Hilfe der Daten versuchen, Rechtsansprüche geltend zu machen.

Nach der Wiedervereinigung der beiden deutschen Staaten im Oktober 1990 wurden ostdeutsche Behörden, die nicht von Bundes- oder Länderbehörden übernommen wurden, entweder privatisiert oder aufgelöst. Viele staatliche Rechenzentren wurden geschlossen und in diesem Auflösungsprozess wurden zahlreiche Datenbestände systematisch zerstört oder in neue private Unternehmen verschoben. Das Bundesarchiv musste sofort eine Übernahmestrategie entwickeln und konnte so eine große Anzahl von historisch und rechtlich relevanten Unterlagen retten. Es war vor allem wichtig, nicht nur die Daten selbst, sondern auch die begleitenden Doku-

3 Dobratz, Susanne; Neuroth, Heike; Schoger, Astrid; Strathmann, Stefan, nestor – Entwicklungsstand des Kompetenznetzwerkes zur Langzeitarchivierung digitaler Ressourcen in Deutschland, in: Zeitschrift für Bibliothekswesen und Bibliographie Jahrgang 52 (2005), S. 151-162, hier S. 152.

mentationsunterlagen zu sichern. Nur mit diesen Unterlagen konnten die Daten interpretiert werden. Die Übernahme war und ist bis heute ein fortdauernder Prozess. Über die Jahre wurden hunderte von Magnetbändern auf Dachböden und in verlassenen Büros aufgefunden. Selbst heute werden noch Disketten in Papierakten im Zuge der Erschließung aufgefunden. Die jüngste Entdeckung sind die Daten eines Untersuchungsausschusses der letzten Volkskammer der DDR. Dieser Untersuchungsausschuss wurde eingesetzt, um Korruptionsvorwürfe gegen die Regierung aufzuklären. Die Daten waren auf 5.25 Zoll Disketten im Format Redabas gespeichert, der sozialistischen Version des Datenbankprogramms dBase.

Einer der wichtigsten Datenbestände ist der Zentrale Kaderdatenspeicher. Dieser enthält Daten, die in den 1980er Jahren im Auftrag des DDR-Ministerrates erfasst wurden. Es sind Informationen über mehr als 700.000 Personen einschließlich der funktionalen Elite und der Entscheidungsträger der DDR (mit Ausnahme der Vollzeit-Parteifunktionäre, der Militärs und der Bediensteten der Staatssicherheit). Diese einzigartige Quelle liefert vollständige Angaben zum sozialen und familiären Hintergrund der hochrangigen Parteimitglieder, zu ihrer Familiensituation in den achtziger Jahren, ihrem beruflichen Werdegang, ihrer Ausbildung, der Zugehörigkeit zu Partei- und Massenorganisationen, ihrer Stellung im Parteiapparat und weitere Informationen bezüglich ihrer Stellung als Kader (wie Fremdsprachenkenntnisse, Berechtigung für Reisen in das nicht-sozialistische Ausland) usw. Diese Daten werden – in anonymisierter Form – schon seit vielen Jahren von einem Sonderforschungsbereich an der Universität Jena ausgewertet, bei dem es um die Erstellung einer kollektiven Biografie der funktionalen Elite der DDR geht.

Andere Beispiele wichtiger Datenbestände sind die Volkszählungsdaten von 1971 und 1981 oder die Daten über die Eingaben der Bürger beim Staatsrat der DDR; welche die Petitionen von mehr als einer Million DDR-Bürgern aus der Zeit von 1979 bis 1989 enthalten.

### Langzeitarchivierungsstrategien

Im Zuge der ersten Datenübernahmen musste auch eine Langzeitarchivierungsstrategie aufgestellt werden. Nach Margret Hedstrom bedeutet Lang-

zeiterhaltung, dass die Fähigkeit erhalten wird, digitale Informationen abzuspielen, wieder zu finden, zu verändern und zu benutzen angesichts sich beständig verändernder Technologien.<sup>4</sup> Für archivische Zwecke geht es nicht nur um die Erhaltung für die kommenden Monate oder Jahre, sondern um die Langzeiterhaltung. Langzeit meint dabei, dass die Zeitspanne lang genug ist, um von den Auswirkungen sich ändernder Technologien, einschließlich neuer Medien oder Formate, oder von einer sich ändernden Nutzergruppe betroffen zu sein. Langzeit kann sich bis in die Unendlichkeit ausdehnen.

Im Allgemeinen gibt es zwei verschiedene Strategien für die Langzeiterhaltung originär digitaler Daten: Migration und Emulation. Migration kann im weitesten Sinne definiert werden als der Transfer digitaler Informationen von einer Hardware oder Software-Plattform zu einer anderen oder von einer Generation der digitalen Technologie zu der nächsten. Migration ist immer ein fortlaufender Prozess. Denn sie kann eigentlich niemals in einem permanent gültigen Datenformat resultieren. Stattdessen müssen Migrationen in regelmäßigen Abständen durchgeführt werden, um sicherzustellen, dass die Informationen interpretierbar bleiben. Emulation als zweite Möglichkeit versucht, die originäre Hard- oder Softwareumgebung zu erhalten. Emulatoren rekonstruieren den so genannten „look and feel“ des Originals, indem sie die Software und Hardware, mit der das Dokument entstanden ist, simulieren. Die Benutzer/innen können somit das Dokument in seiner ursprünglichen Form mit allen Funktionalitäten nutzen. Aus archivwissenschaftlicher Perspektive ist dies ohne Zweifel der beste Weg, weil er soviel vom Original beibehält wie möglich. Jedoch ist es auch der komplizierteste und teuerste Weg.

Die Wahl der Strategie hängt schließlich von dem Grad der Funktionalität, die erhalten werden soll, und von den Ressourcen, die ein Archiv in die digitale Langzeiterhaltung investieren kann, ab. Man kann heute feststellen, dass sich keine der beiden Strategien als die eindeutig bessere

4 Hedstrom, Margaret, Preserving Digital Information, in: Long Term Preservation of Electronic Materials. A JISC/British Library Workshop as part of the Electronic Libraries Programme (eLib), vgl. <<http://www.ukoln.ac.uk/services/papers/bl/rdr6238/paper.html#hedstrom>> (16.02.2006).

Lösung darstellt. Langzeiterhaltung ist und bleibt ein technisch komplexer und kostenintensiver Prozess. Das vorrangige Ziel sollte sein, die Daten für alle möglichen Fragen offen zu halten, die im Lauf der Zeit gestellt werden könnten. Langfristig können neue Benutzergruppen entstehen, deren Erwartungen sich deutlich von denen heutiger Nutzer/innen unterscheiden. Migration und Emulation bedeuten beide fortlaufende Verpflichtungen bei der Langzeiterhaltung – ein Archiv muss in der Lage sein, Ressourcen in regelmäßigen Abständen bereitzustellen, um die Benutzbarkeit der Daten zu sichern. Selbst wenn man die Fragen der logischen Interpretierbarkeit der Daten außer Acht lässt, sind die physischen Speichermedien dem Verfall ausgesetzt und müssen regelmäßig erneuert werden.

### **Migration im Bundesarchiv**

Das Bundesarchiv hat sich bei den DDR-Daten für eine Migrationsstrategie entschieden. Als Langzeiterhaltungsformat wurde ASCII gewählt. ASCII ist der sicherste Weg, um Daten langfristig zu speichern, aber es wird auch lediglich der Bitstream ohne jede Information über die ursprünglichen Funktionalitäten bewahrt. Diese Information muss durch zusätzliches Dokumentationsmaterial gesichert werden. In vielen Fällen mussten bei den DDR-Daten die Datenstrukturen rekonstruiert werden. Viele Recherchen waren erforderlich, um Datensatzstrukturen und Kodierungen wieder herzustellen. Um Speicherplatz zu sparen, der in den 1970er und 1980er Jahren sehr teuer war, wurden die Informationen so weit es ging kodiert, gepackt oder komprimiert. Die Konvertierung in ein langzeitarchivierungsfähiges Format konnte daher nicht mit marktüblichen Konvertierungsprogrammen durchgeführt werden. Jeder Datenbestand erforderte die Programmierung spezieller Konvertierungsprogramme. In einigen Fällen mussten die Programmierer/innen ausfindig gemacht werden, welche die Komprimierungsalgorithmen geschrieben hatten, weil sie die einzigen waren, die die Komprimierungen auflösen konnten.

2003 hat das Bundesarchiv zusammen mit der Universität Koblenz ein Programm entwickelt, das es erlaubt, den Großteil der Datenbestände von EBCDIC nach ASCII oder auch XML zu konvertieren. Das Programm löst die gepackten und komprimierten Daten auf und konvertiert hexadezimale

in binäre Werte. Mit diesem Programm können die einzelnen Feldinhalte analysiert werden, so dass Datensatzstrukturen und Codierungen validiert werden können. Die wichtigste Funktion des Programms besteht darin, dass Daten, Datensatzstruktur und Kodierungen untrennbar miteinander verbunden werden können. Diese Funktion kann den Nachteil von ASCII, das lediglich die Speicherung des Bitstreams erlaubt, ausgleichen.

### Vertrauenswürdigkeit, Authentizität – Metadaten

Die Bearbeitung der Daten wurde bisher in speziellen Findmitteln dokumentiert. Dazu gehören auch Informationen über Fehler oder mögliche Vorbehalte gegenüber der Zuverlässigkeit der Daten ebenso wie die Informationen über die Original Soft- und Hardware, die technische Bearbeitung usw. Im Moment erprobt das Bundesarchiv im Rahmen von nestor die Umsetzung dieser Informationen in internationale Metadatenstandards.

Die Dokumentation ist zwingend notwendig, um die Authentizität der Quellen langfristig zu wahren. Die Frage, wie die Authentizität elektronischer Unterlagen erhalten und nachgewiesen werden kann, ist eine der Kernfragen nicht nur für die Archivare/innen, sondern für die gesamte Informationsgesellschaft. Grundlegend ist somit der Begriff der Authentizität. Was macht ein Objekt zu einem authentischen Objekt? Cullen bemerkt hierzu: „An authentic object is one whose integrity is intact – one that is and can be proven or accepted to be what its owners say it is. It matters little whether the object is handwritten, printed, or in digital form.“<sup>5</sup>

Es gibt vielseitige Möglichkeiten, um die Authentizität eines analogen Objekts nachzuweisen. Am bekanntesten sind sicher die chemischen Analysen des Papiers oder der verwendeten Tinte. Man kann aus der chemischen Analyse das Herstellungsverfahren des Papiers oder der Tinte ableiten. Diese Herstellungsverfahren lassen sich einem Zeitraum zuordnen. Stimmt dieser nicht mit dem auf dem Dokument angegebenen Datum

5 Cullen, Charles T., Authentication of Digital Objects. Lessons from a Historian's Research, in: Authenticity in a Digital Environment, Washington, D.C. 2000, S. 1-7, hier S. 1, vgl. <<http://www.clir.org/pubs/abstract/pub92abst.html>> (16.2.2006).

überein, so sind Zweifel an der Authentizität des Dokuments angebracht. Dies ist nur eine Möglichkeit einer Authentizitätsprüfung eines analogen Objekts. Dass Information und Informationsträger, in den meisten Fällen Papier, eine untrennbare Verbindung eingegangen sind, ist hierbei ein großer Vorteil. Papier, Tinte und Schriftbild sowie der Allgemeinzustand sind Teil des intrinsischen Wertes eines Objekts und somit mit ihm verbunden.

Auch digitale Objekte müssen auf einem Datenträger, einem Medium, gespeichert werden. Im Gegensatz zum analogen Objekt ist die Wahl des Datenträgers jedoch unabhängig von der Art des Objekts. Während in der analogen Welt das Papier beinahe immer die preisgünstigste und einfachste Alternative darstellt, kann man in der digitalen Welt jede technisch mögliche Lösung auch scheinbar problemlos anwenden. Dieser Umstand und die zusätzliche Möglichkeit zur schnellen Datenübertragung macht die Bindung der Information an ihren Träger extrem flüchtig. Auch andere Kriterien zur Authentizität aus der analogen Welt bekommen das digitale Objekt nicht in den Griff. Deshalb muss man zunächst das „Wesen“ eines digitalen Objekts verstehen, bevor man sich auf die Suche nach seiner Authentizität macht. Die besondere Eigenschaft eines digitalen Objektes ist seine „Existenz“ auf drei Ebenen. Jedes digitale Objekt kann gleichzeitig als ein physisches, logisches und konzeptuelles Objekt betrachtet werden.<sup>6</sup>

- *physisches Objekt*: Zeichen, die auf einem Informationsträger eingeschrieben sind. Die Art und Weise der physischen Beschaffenheit dieser Zeichen kann aufgrund der unterschiedlichen Beschaffenheit des Trägers sehr unterschiedlich sein. Auf einer CD-ROM sind es die sogenannten „pits“ und „lands“ auf der Trägeroberfläche, bei magnetischen Datenträgern sind es Übergänge zwischen magnetisierten und nicht magnetisierten Teilchen (die sogenannten Flusswechsel). Auf der physischen Ebene haben die Bits keinerlei Bedeutung, das heißt, Bits die zu einem Text gehören unterscheiden sich in keinsten Weise von

6 Thibodeau, Kenneth, Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years, in: The State of Digital Preservation. An International Perspective, Washington D.C. 2002, S. 4-31, hier S. 6-10, vgl. <<http://www.clir.org/pubs/abstract/pub107abst.html>> (16.02.2006).

Bits, die Teil eines Computerprogramms sind.

- *logisches Objekt*: Bits, die vom Informationsträger gelesen werden und nun von einer Software interpretiert werden. In dieser Ebene existieren die Objekte als Bitstrom, der als Dateiformat erkannt wird.
- *konzeptuelles Objekt*: Zeichen und Objekte, die vom Menschen wahrgenommen und sinnvoll interpretiert werden können. Das konzeptuelle Objekt ist die eigentliche, für die Betrachter/innen bedeutungsvolle Einheit (zum Beispiel ein Text, Musikstück, Film, Computerprogramm).

Ausgehend von dieser Grundeigenart, die für alle digitalen Objekte gilt, sind Verfahren, die die Authentizität eines digitalen Objekts anhand einer Autopsie überprüfen wollen, ungeeignet. Der Datenträger kann bereits tausendfach gewechselt haben. Selbst innerhalb eines Archivs wird das digitale Objekt den Träger noch mehrfach wechseln müssen. Seine Prüfung führt deshalb zu keinem Ergebnis. Auch das logische Objekt kann sich im Rahmen einer Datenmigration ohne eine augenscheinliche Veränderung des konzeptuellen Objekts geändert haben. Es kann aber auch der Fall sein, dass eben jene Datenmigration auch das konzeptuelle Objekt gravierend verändert hat, so dass beispielsweise bei einem E-Journal bestimmte Artikel nicht mehr lesbar sind. Wenn die Kriterien für die Authentizität eines Objekts nicht mehr am Objekt selbst abprüfbar sind, wo dann?

Da ein digitales Objekt ständig kopiert und verändert werden kann, müssen in einem Archiv die Veränderungen festgehalten und dokumentiert werden. Wesentliche Fragen zur Authentizität eines digitalen Objekts sind zum Beispiel

- Wer hat wann in welcher Art und Weise das Objekt verändert?
- In welchen Dateiformaten lag und liegt das Objekt vor?
- Sind die Eigenschaften des Objekts verändert worden?
- Aus welcher technischen Umgebung stammt das Objekt ursprünglich?
- Ist das Objekt manipuliert worden?

Diese Daten über digitale Objekte nennt man Metadaten. Sie enthalten Informationen über Informationsobjekte, ähnlich wie bibliografische Beschreibungen, die Informationen über Bücher enthalten. Geordnet werden Metadaten im Allgemeinen in einem Metadatenschema. Es gibt

heute viele Metadaten-Schemata, die für unterschiedliche Zwecke geschaffen wurden. Das Metadaten-Schema Dublin Core beschreibt größtenteils das konzeptuelle Objekt, besitzt aber auch Elemente, die das physische und das logische Objekt mit einbeziehen. Das Schema wurde zu einem besonders großen Anteil von Bibliotheken entwickelt, um Informationen zu finden, zu verwalten und unter Partnern zu teilen. Im Archivbereich ist das Encoded Archival Description (EAD) Metadaten-Schema zur Erstellung von Findmitteln sehr gebräuchlich. Bislang fehlte aber noch ein Metadaten-Schema, das speziell auf die Bedürfnisse der digitalen Langzeitarchivierung ausgerichtet ist. An diese Stelle tritt nun das Metadaten-Schema PREMIS.

### PREMIS

Das Akronym steht für Preservation Metadata Implementation. PREMIS wurde 2003 mit Unterstützung der Research Libraries Group (RLG) und des Online Computer Library Center (OCLC) gegründet. Die Initiative ist ein internationaler Verbund von Archiven, Bibliotheken, Museen und Vertretern aus der privaten Wirtschaft. Das PREMIS Metadaten-Schema ist das Arbeitsergebnis einer kooperativen Struktur. Der Abschlussbericht der Arbeitsgruppe wurde im Mai 2005 veröffentlicht. Die Ziele der Gruppe waren:

- Die Definierung eines implementierbaren Kernsatzes von Metadaten zur digitalen Langzeitarchivierung;
- Die Erstellung eines Data Dictionaries zur Erklärung des Metadaten-Schemas;
- Die Evaluierung alternativer Strategien zur Formatwahl, Speicherung und Verwaltung von Langzeitarchivierungsmetadaten innerhalb eines digitalen Archivs und zum Austausch zwischen verschiedenen Archivsystemen;
- Die Organisation von Pilot Programmen zum Test der Empfehlungen unter unterschiedlichen Rahmenbedingungen;
- Die Erforschung von Möglichkeiten zum Erstellen und Teilen von Langzeitarchivierungsmetadaten.
- Die Arbeit am PREMIS Metadaten-Schema ist noch nicht abgeschlossen.

sen. Seit der Veröffentlichung durchläuft es eine geplante achtzehnmonatige Test- und Implementierungsphase.<sup>7</sup>

PREMIS ist ein geeignetes Werkzeug, um die Authentizität eines digitalen Objekts zu überprüfen und über einen langen Zeitraum auch zu erhalten. Die gespeicherten Metadaten geben Auskunft über das archivierte digitale Objekt. Natürlich können die Informationen, die in einem Metadatenschema gespeichert und geordnet werden, auch in anderer Form dargestellt werden. Bislang wurden sie im Bundesarchiv in den Dossiers in natürlicher Sprache auf Papier ausgedruckt, sie liegen aber auch als Text in elektronischer Form vor. Das folgende Beispiel beschreibt den Verlauf einer typischen Migration einer Datentabelle aus dem Bestand der DDR Volkszählungsdaten der Jahre 1971 und 1981. Der Text gibt Auskunft auf die Frage: „Wer hat wann in welcher Art und Weise das Objekt Datei 0009\_A9881.B7101.N11 verändert?“

Die Datei mit dem Namen 0009\_A9881.B7101.N11 wurde am 20.10.2005 um 15:00 Uhr mitteleuropäischer Zeit vom Programm „Konverter.exe“ konvertiert. Diese Aktion ist nach Wunsch verlaufen. Die Datei 0009\_A9881.B7101.N11 liegt im EBCDIC Format vor. Nach der Migration sind zwei neue Dateien mit dem gleichen Informationsinhalt entstanden. Die beiden neuen Dateien heißen 0009\_A9881.B7101.N11\_1\_0.fed und 0009\_A9881.B7101.N11\_2\_0.fed und liegen im ASCII.csv Format vor.

Ein Vorteil des Dossiers ist seine leichte Lesbarkeit für das menschliche Auge. Der große Nachteil indes liegt in der schlechten maschinellen Verwertbarkeit. Einen reinen Text kann man nicht strukturiert in eine relationale Datenbank überführen. Zudem ist die Semantik der einzelnen Begriffe nicht maschinell erschlossen. Hier setzt nun ein Metadatenschema an. Es ist zwar für den Menschen auf den ersten Blick viel schwerer lesbar, in ihm wird die Information jedoch in immer gleicher Art und Weise strukturiert und semantisch erschlossen. Dadurch lassen sich die Informationen ma-

7 Data Dictionary for Preservation Metadata. Final Report of the PREMIS Working Group, Dublin, Ohio 2005 hier S. vii, vgl. <<http://www.oclc.org/research/projects/pmwg/premis-final.pdf>> (16.02.2006).

schinell verarbeiten. PREMIS nutzt die Auszeichnungssprache XML als Format zur Speicherung und Vermittlung der Metadaten. Die Strukturen und Datentypen von PREMIS wurden innerhalb von XML-Schema Dateien definiert, welche zur freien Verfügung bereitstehen. Diese Schema-Dateien bilden die Grundlage für die Tests am Bundesarchiv. Sie werden zur Zeit sukzessive an die Bedürfnisse eines Archivs angepasst.

Der obige Text verteilt sich auf die folgenden PREMIS-Elemente:

```
<eventType>Migration</eventType>
<eventDateTime>2005-10-20T15:00:00
+01:00</eventDateTime>
<eventDetail>EBCDIC Datei wird in zwei ASCII csv
Dateien migriert. Dateiendungen aus technischen
Gründen fed.</eventDetail>
<eventOutcomeInformation>
<eventOutcome>OK</eventOutcome>
</eventOutcomeInformation>
```

Die Elemente sind bereits im XML-Format dargestellt. Die Elementnamen sind in spitze Klammern gesetzt. Die Metadaten werden durch zwei so genannte Tags eingerahmt, die jeweils den entsprechenden Elementnamen beinhalten, wobei der zweite Tag (End-Tag) mit einem Schrägstrich beginnt. Dem Element mit dem Namen „eventType“ (Ereignis Typ) wurde hier der Begriff Migration zugewiesen. Das Element eventDateTime gibt den genauen Zeitpunkt des Ereignisses wieder. Unter „eventDetail“ wird der Vorgang näher beschrieben und das Element „eventOutcome“ sagt uns, dass die Migration technisch korrekt verlaufen ist. Somit wurde dokumentiert, welche Veränderung am digitalen Objekt vorgenommen wurde, wann die Veränderung durchgeführt wurde und ob die Prozedur nach Plan verlaufen ist.

Die folgenden Elemente verweisen auf das Programm, mit dem die Migration durchgeführt wurde, sowie auf seine Rolle innerhalb des Migrationvorgangs. Das Programm diente als Konvertierungsprogramm, das heißt, es war verantwortlich für die Übersetzung der Ursprungsdatei in ein anderes Dateiformat.

```
<linkingAgentIdentifier>
<linkingAgentIdentifierType>Barch_ID_B2_Agent-
```

```

    nameschlüssel </linkingAgentIdentifizierType>
<linkingAgentIdentifizierValue>ans_konverterEXE
  </linkingAgentIdentifizierValue>
<linkingAgentRole>Konvertierer</linkingAgentRole>
</linkingAgentIdentifizier>

```

Die folgenden Zeilen beschreiben das Konvertierungsprogramm näher. Sie enthalten die Information, dass es sich um ein Programm mit dem Namen Konverter.exe handelt.

```

<agentIdentifizier>
<agentIdentifizier-
  Type>Barch_ID_B2_Agentnameschlüssel</
  agentIdentifizierType>
<agentIdentifizierValue>ans_konverterEXE</agent
  IdentifizierValue>
</agentIdentifizier>
<agentName>Konverter.exe</agentName>
<agentType>Programm</agentType>

```

Alle bisher genannten Informationen werden über Schlüssel (linkingObjectIdentifizier) mit den Informationen über die von der Migration betroffenen Dateien verknüpft. In diesem Beispiel sind drei Dateien von der Migration betroffen. Aus der Quelldatei fns\_0009\_A9881.B7101.N11 entstehen die Dateien 0009\_A9881.B7101.N11\_1\_0.fed und 0009\_A9881.B7101.N11\_2\_0.fed im ASCII/CSV Format.

```

<linkingObjectIdentifizier>
<linkingObjectIdentifizierType>Barch_ID_B2_file-
  nameschlüssel</linkingObjectIdentifizierType>
<linkingObjectIdentifizierValue>fns_0009_A9881.
  B7101.N11</linkingObjectIdentifizierValue>
</linkingObjectIdentifizier>
<linkingObjectIdentifizier>
<linkingObjectIdentifizier-
  Type>Barch_ID_B2_filenameschlüssel</linking
  ObjectIdentifizierType>
<linkingObjectIdentifizierValue>0009_A9881.B7101.
  N11_1_0.fed</linkingObjectIdentifizierValue>
</linkingObjectIdentifizier>
<linkingObjectIdentifizier>
<linkingObjectIdentifizierType>Barch_ID_B2_file-

```

```

    nameschlüssel</linkingObjectIdentifizierType>
<linkingObjectIdentifizierValue>0009_A9881.B7101.
  N11_2_0.fed</linkingObjectIdentifizierValue>
</linkingObjectIdentifizier>

```

Bei den Metadaten zu den jeweiligen Dateien findet man Informationen über die Dateigröße in Kilo Byte und das entsprechende Dateiformat. Das Element „formatRegistry“ gibt einem die Möglichkeit, auf ein Formatverzeichnis außerhalb des Archivs zu verweisen. Innerhalb eines solchen Verzeichnisses werden die unterschiedlichen Dateiformate technisch ausführlich beschrieben und ihre Datenstruktur wird definiert. Zur Zeit sind einige Formatverzeichnisse in Planung. In Harvard wurde gerade ein Projekt begonnen, aus dem einmal ein globales Verzeichnis entstehen soll.<sup>8</sup> Das Beispiel verweist auf das Verzeichnis PRONOM<sup>9</sup> des britischen Nationalarchivs. Hinter dem Schlüssel „id 45“ verbirgt sich die technische Beschreibung des Dateiformats ASCII/CSV.

```

<size>72471552</size>
<format>
<formatDesignation>
<formatName>ascii/csv</formatName>
<formatVersion>1.0</formatVersion>
</formatDesignation>
<formatRegistry>
<formatRegistryName>PRONOM </formatRegistryName>
<formatRegistryKey>id 45</formatRegistryKey>
</formatRegistry>
</format>

```

Veränderungen und Manipulationen an Dateien können über Hash-Werte erkannt werden. Dabei wird eine spezielle Hash-Funktion über dem Bitstrom einer Datei ausgeführt. Dadurch wird ein typischer Hash-Wert erzeugt. Dieser wird üblicherweise durch eine Zeichenkette im Hexadezimalcode dargestellt, der als digitaler Fingerabdruck verwendet werden kann. Bereits geringfügigste Veränderungen an der Datei durch Manipula-

8 Global Digital Format Registry (GDFR), vgl. <<http://hul.harvard.edu/gdfr/about.html>> (16.2.2006).

9 The technical registry PRONOM, vgl. <<http://www.nationalarchives.gov.uk/pronom/>> (16.02.2006).

tion oder fehlerhafte Datenübertragung führen zu einem völlig veränderten Hash-Wert. Da es verschiedene Hash-Funktionen gibt, muss der Name der Funktion ebenso bei den Metadaten gespeichert werden, wie der entsprechende Hash-Wert. PREMIS hat die entsprechenden Elemente dafür. Die Hash-Funktion des Beispiels hat den Namen MD5. Der entstandene Wert hat im Hexadezimalcode den Wert „7c9b35da4f2ebd436f“. Die entsprechende Integritätsprüfung wurde vor der Abgabe der Datei an das Bundesarchiv im Statistischen Bundesamt durchgeführt.

```
<fixity>
<messageDigestAlgorithm>MD5</messageDigest-
  Algorithm>
<messageDigest>7c9b35da4f2ebd436f</messageDigest>
<messageDigestOriginator>StatistischesBundesamt
  </messageDigestOriginator>
</fixity>
```

Die vorgestellten Elemente sind nur ein kleiner Ausschnitt aus dem Metadatenschema PREMIS. Dennoch zeigen sie, wie die Authentizität von digitalen Objekten innerhalb eines Archivs überprüfbar bleibt, obwohl das Objekt gravierende Veränderungen erfährt. Festzuhalten ist:

- Wir können nachvollziehen, inwiefern ein Objekt mit der Zeit verändert wurde und welche Konsequenzen dies für seinen Informationswert hat.
- Wir können nachvollziehen, wann und warum diese Änderungen vorgenommen wurden und wer dafür verantwortlich ist.
- Wir können durch spezielle Funktionen überprüfen, ob die Objekte ungewollt verändert wurden.
- Wir können die digitalen Objekte technisch umfassend beschreiben.

### Bisheriger Umgang mit Authentizität

Authentizität ist kein Selbstzweck. Wie bereits erwähnt, werden die Datenbestände des Bundesarchivs häufig zu Nachweiszwecken benutzt. Bürger/innen der ehemaligen DDR nutzen die Daten, um ihre Ansprüche auf Wiedergutmachung, Entschädigungen, Rentenzahlungen durchzusetzen. Einer der wichtigsten Datenbestände ist zum Beispiel der Datenbe-

stand „Datenspeicher gesellschaftliches Arbeitsvermögen“, der Informationen zu einem Großteil der Beschäftigten in der ehemaligen DDR enthält, einschließlich der Informationen über Ausbildung und Beschäftigungsverhältnisse von mehr als 7,25 Millionen Bürger/innen. Diese Quelle dient oft als Nachweis von Beschäftigungsnachweisen für Rentenansprüche. Die größte Anzahl von Anfragen bezieht sich auf den Datenbestand „Strafgefangenen- und Verhaftetendaten“. Diese Daten wurden im Auftrag des Innenministeriums der DDR erhoben. Sie enthalten Informationen über die familiäre Situation, die Haftzeit, die Straftat und über alle Ereignisse, die sich während der Haftzeit zugetragen haben, unter anderem über die Krankheiten.

Besonders die Nachweise über während der Haft erlittene Krankheiten oder Unfälle können ehemaligen Häftlingen heute dabei helfen, Entschädigungen zu bekommen. Diese Informationen gibt es nur in elektronischer Form und es ist die einzige Quelle, auf die sich Betroffene berufen können, wenn sie eine Entschädigung beantragen. In erster Instanz richten sie ihre Forderungen an die lokalen Behörden, die über den Antrag entscheiden. Wenn diese den Antrag ablehnen, besteht die Möglichkeit, den Anspruch vor Gericht einzuklagen, was häufig vorkommt. Wie gezeigt, entsprach die Übernahme der Daten in das Bundesarchiv in den 1990er Jahren keineswegs den Regeln eines ordnungsgemäßen, sicheren Datentransfers. Aber keine Behörde und kein Richter, keine Richterin hat bis heute die Authentizität der vorgelegten Daten angezweifelt. Scheinbar hat die sorgfältige Dokumentation der Übernahme und Bearbeitung der Daten Behörden und Richter/innen zufrieden gestellt.

### Fazit und Ausblicke

Diese Erfahrung bestätigt die Politik des Bundesarchivs, anstelle von kurz- oder mittelfristigen Lösungen für die technische Sicherung der Authentizität, wie zum Beispiel durch elektronische Signaturen, auf organisatorische Lösungen für die Langzeitsicherung zu setzen. Das Bundesarchiv arbeitet aktiv bei der nestor Arbeitsgruppe „Vertrauenswürdige Archive – Zertifizierung“ mit. Hierbei handelt es sich um ein interdisziplinäres Projekt, das nach dem Vorbild der Research Library Group Kriterien für eine Zertifizie-

rung vertrauenswürdiger Archive entwickelt. Das Ziel des Projektes besteht darin, Anforderungen für die vertrauenswürdige und damit authentische Speicherung, Bearbeitung und Nutzbarmachung elektronischen Archivguts aufzustellen. Diese Kriterien kann keine Institution allein aufstellen, dazu bedarf es einer breiten Kooperation, wie sie mit nestor gegeben ist.

Zusammenfassend können wir sagen, dass die Wahrung der Authentizität von digitalen Objekten grundlegend für eine zukünftige Überlieferungsbildung ist. Vertrauenswürdige Archive, die in der Lage sind, digitale Objekte authentisch zu speichern, werden in Zukunft an Bedeutsamkeit zunehmen. Kooperative Strukturen wie nestor, die Kompetenzen und Material aus verschiedenen Bereichen zusammenbringen, sind wichtig zur Erforschung und Diskussion von neuen Kriterien und Techniken in digitalen Archiven. Sie tragen somit ihren Teil zu einer gesicherten geschichtlichen Überlieferung bei.

\*\*\*

*Dr. Andrea Hänger arbeitet seit 2000 am Bundesarchiv in Koblenz. Dort leitet sie das Referat für elektronische Archivierung. Sie arbeitet in zahlreichen nationalen und internationalen Gremien zur digitalen Langzeitarchivierung mit, unter anderem vertritt sie das Bundesarchiv als Partner im Projekt nestor, war Leiterin der Arbeitsgruppe, die das DOMEA-Aussonderungskonzept erstellt hat, und ist Vorstandsmitglied im europäischen DLM Forum. E-Mail: a.haenger@barch.bund.de*

*Karsten Huth arbeitet im Rahmen des Projekts nestor am Bundesarchiv in Koblenz. Zur Zeit entwickelt er ein Metadatenchema zur digitalen Langzeitarchivierung. E-Mail: k.huth@barch.bund.de*

## LANGFRISTIGE VERFÜGBARKEIT VON QUELLENEDITIONEN IM DIGITALEN ZEITALTER – TEXT ENCODING INITIATIVE, GESCHICHTSWISSENSCHAFT UND BIBLIOTHEKEN

von *Stefan Cramme*

*Editionen gehören zum grundlegenden Handwerkszeug der historischen Wissenschaften. Aus verschiedenen Gründen ist für sie in den letzten Jahren die elektronische Publikation attraktiv geworden. Zur Gewährleistung der langfristigen Verfügbarkeit bietet sich die Verwendung der Richtlinien der Text Encoding Initiative (TEI) an, die auf offenen Standards basieren und unabhängig von möglichen Ausgabeformaten sind. In diesem Zusammenhang kann auch Bibliotheken eine große Rolle zukommen, die als Dienstleister bei der Erstellung und dauerhaften Archivierung digital aufbereiteter Quellenbestände mitwirken können. Vorgestellt werden die Erfahrungen der Bibliothek für Bildungsgeschichtliche Forschung mit der TEI-Auszeichnung für Werkausgaben von Adolf Reichwein und Briefe Friedrich Fröbels.*

\*\*\*

### **Historische Quellen im digitalen Zeitalter**

Texteditionen stellen nicht nur in den philologischen Disziplinen, sondern auch in den historisch arbeitenden Wissenschaften nach wie vor eine unverzichtbare Arbeitsgrundlage dar. Sie gewährleisten insbesondere die dauerhafte Verfügbarkeit und Überprüfbarkeit schriftlicher Quellen<sup>1</sup>,

<sup>1</sup> Die Begriffe Quelle und Edition, ihr Sinn oder möglicherweise auch Unsinn für die verschiedenen Bereiche der Geschichtswissenschaft können hier nicht weiter reflektiert werden. Siehe etwa Gall, Lothar; Schieffer, Rudolf (Hgg.), Quelleneditionen und kein Ende? (Historische Zeitschrift, Beihefte 28), München 1999;

unabhängig von der Möglichkeit, eventuell weit entfernte Archive oder Bibliotheken aufsuchen zu müssen, oder der Gefahr des Verlustes des Originals.<sup>2</sup> Die jahrhundertlang bewährte Veröffentlichung in Form gedruckter Bände stößt aber in letzter Zeit auf vor allem finanzielle Schwierigkeiten, weil Verlage angesichts der schlechten Absatzmöglichkeiten mit Recht sehr zurückhaltend sind, insbesondere bei umfangreichen und über einen langen Zeitraum geplanten Editionsprojekten, wie sie oftmals üblich waren.

In den letzten Jahren hat sich für Editionen die elektronische Erfassung und Bearbeitung der Texte durchgesetzt, zunächst als Vorstufe einer traditionellen Druckausgabe. Zunehmend tritt allerdings auch die direkte Bereitstellung elektronischer Fassungen für die Nutzenden ins Blickfeld.<sup>3</sup> Solche elektronischen Editionen<sup>4</sup> bieten gegenüber gedruckten Bänden klar

---

Quellen und Quellenedition im neuen Medienzeitalter (Fundus – Forum für Geschichte und ihre Quellen 2), unter <[http://webdoc.sub.gwdg.de/edoc/p/fundus/html/heft\\_2.html](http://webdoc.sub.gwdg.de/edoc/p/fundus/html/heft_2.html)> (alle Links wurden zuletzt überprüft am 26.09.2006); Fellner, Fritz, Die historische Quelle – Instrument der Geschichtsforschung und Baustein des Geschichtsbewußtseins oder Baustein der Geschichtsforschung und Instrument des Geschichtsbewußtseins?, in: Klingenstein, Grete; Fellner, Fritz; Hye, Hans Peter (Hgg.), Umgang mit Quellen heute. Zur Problematik neuerzeitlicher Quelleneditionen vom 16. Jahrhundert bis zur Gegenwart (Fontes rerum Austriacarum, 2. Abteilung Diplomataria et acta 92), Wien 2003, S. 19-33. Zur literaturwissenschaftlichen Editionsphilologie: Nutt-Kofoth, Rüdiger (Hg.), Text und Edition. Positionen und Perspektiven, Berlin 2000; Kraft, Herbert, Editionsphilologie, 2. neubearb. Aufl., Frankfurt am Main 2001. Alle in den Anmerkungen genannten URLs wurden zuletzt am 28. August 2006 überprüft.

2 Zur elektronischen Bereitstellung von nicht editorisch bearbeitetem Archivmaterial siehe beispielsweise Jucker-Kupper, Patrick; Koller, Christophe; Ritter, Gerold (Hgg.), Digitales Gedächtnis. Archivierung und die Arbeit der Historiker der Zukunft (Geschichte und Informatik 13/14), Zürich 2004.

3 Eine aktuelle Übersicht über geschichtswissenschaftliche Editionsprojekte ist nur schwer zu erhalten. Eine auf Österreich bezogene Umfrage 1999/2000 ergab, dass weniger als 10 Prozent eine elektronische Edition anstrebten (Klingenstein u.a. (Hgg.), Umgang (wie Anm. 1), S. 211-325; kurze Auswertung: Ehmer, Josef; Steidl, Annemarie, Überlegungen zur digitalen Edition historischer Quellen und Vorstellung des Editionsprojekts Wiener Innungsarchivalien auf CD-ROM, ebd. S. 191-192). Ankündigungen neu begonnener Projekte weisen jedoch in den letzten Jahren regelmäßig zumindest auf die Möglichkeit digitaler Publikationsformen hin.

4 Zur noch unscharfen Terminologie: Kamzelak, Roland, Hypermedia – brauchen wir eine neue Editionswissenschaft?, in: Ders. (Hg.), Computergestützte Text-

erkennbare Vorteile:

- Suchmöglichkeiten im Volltext, insbesondere, wenn sie von der Auszeichnung bestimmter Elemente (Namen, Schlagwörter, Sprachliches usw.) unterstützt werden;
- in digitaler Form können prinzipiell größere Textmengen verarbeitet werden als bei gedruckten Bänden;
- Herstellen von Bezügen und Verweisen innerhalb des Textes, aber auch auf andere elektronisch vorliegende Texte, auf Kommentare oder ergänzendes Material<sup>5</sup>; die Benutzeroberfläche kann dabei, muss aber nicht, vom bisher Gewohnten abweichen<sup>6</sup>;
- Weiterverarbeitung, insbesondere Umwandlung in verschiedene Ausgabeformen;
- leichtere Publikation von Zwischenergebnisse und korrigierten/überarbeiteten Fassungen. (Dies ist gerade bei Editionen, die eine zitierfähige, stabile Textfassung gewährleisten sollen, natürlich nicht unproblematisch.)

Diesen unverkennbaren Vorteilen stehen allerdings auch gewichtige Einwände gegenüber:

- das Rezeptionsverhalten zumal bei längeren Texten ist immer noch auf das Lesen einer gedruckten Fassung ausgerichtet<sup>7</sup>;
- die dauerhafte Verfügbarkeit elektronischer Dokumente ist durch die ständige Weiterentwicklung von Soft- und Hardware gefährdet;
- insbesondere in der Übergangs- und damit auch Experimentierphase sind keine zeitlichen und finanziellen Einsparungen zu erwarten.

Diesen Nachteilen kann mit einer Doppelstrategie begegnet werden:

---

Edition (Beihefte zu Editio 12), Tübingen 1999, S. 119-126, hier S. 120; zur Abgrenzung von nicht editierten elektronischen Texten im Internet: Pianos, Tamara, Das Canterbury Tales Project, in: Bibliothek und Wissenschaft 36 (2003), S. 95-150, hier S. 110.

5 Siehe Eibl, Karl; Jannidis, Fotis; Willems, Marianne, Der Junge Goethe in neuer Ausgabe. Einige Präliminarien und Marginalien, in: Kamzelak (Hg.), Text-Edition (wie Anm. 4), S. 69-78, hier S. 73-74.

6 Jenks, Stuart, Dekonstruktion und Rekonstruktion der Quellenedition, in: Mitteilungen aus dem Bundesarchiv 11 (2003), Heft 1, S. 5-13.

7 Siehe Eibl u.a., Junge Goethe (wie Anm. 5), S. 71-72.

- Hybridausgaben, die gedruckte und elektronische Fassung kombinieren, verbinden die Vorteile beider Medienarten;
- zur Auszeichnung und Speicherung der elektronischen Fassungen ist auf offene und standardisierte Formate zurückzugreifen.

### Offene Formate und Auszeichnungssprachen

Die Verwendung proprietärer, also von einem Hersteller oder einem bestimmten Softwareprodukt verwendeter, nicht offen dokumentierter und anpassbarer Formate ist unter dem Gesichtspunkt der langfristigen Verfügbarkeit auf jeden Fall zu vermeiden und durch echte, das heißt offen gelegte und breit unterstützte Formate zu ersetzen, die unabhängig von bestimmter Anwendungssoftware oder Betriebssystemen genutzt werden können. Eine aktuell weite Verbreitung kann dabei kein Gegenargument sein, weil etwa das Dateiformat der Microsoft-Office-Programme mehrfach geändert worden ist und erst jetzt unter dem Druck von Mitbewerbern eine zögerliche Offenlegung der Formate erfolgt, aber nur für zukünftige Versionen. (Bei Multimediaformaten sind die möglichen Gefahren noch größer.<sup>8</sup>)

Etwas besser sieht es bei dem von Adobe eingeführten Portable Document Format (PDF) aus, das zwar kein offener Standard im eigentlichen Sinne ist, aber zumindest von einer Vielzahl von Anwendungen plattformübergreifend geschrieben und gelesen werden kann. Der Nachteil von PDF liegt aber darin, dass es vor allem das Erscheinungsbild eines Dokuments beschreibt, nicht die zugrunde liegenden logischen Strukturen. PDF ist ein Endformat, das sich nicht sinnvoll weiterverarbeiten lässt. Für Texte ist daher die Verwendung von Auszeichnungssprachen anzuraten, die eine Trennung der inhaltlichen Struktur von der Erscheinungsform ermöglichen. Schon seit den 1980er Jahren gibt es SGML (Standard General

8 Beispielhaft sei auf das Projekt der BBC für eine digitale Version des Domesday Books hingewiesen, das keine zwei Jahrzehnte nach seiner Erstellung nur noch unter Verwendung musealer Hardware zu verwenden war: Darlington, Jeffrey; Finney, Andy; Pearce, Adrian, Domesday redux. The rescue of the BBC Domesday Project videodiscs, in: Ariadne 36 (2003), unter <<http://www.ariadne.ac.uk/issue36/tna/>>.

Markup Language) als international standardisiertes Meta-Format zur Definition von Auszeichnungssprachen. Es hat weite Verbreitung vor allem beim technischen Publizieren gefunden. Als Weiterentwicklung, teilweise auch Vereinfachung, in spezieller, aber nicht ausschließlicher Hinsicht auf die Anwendung im World Wide Web (WWW) ist jetzt vor allem XML (Extensible Markup Language) zu nennen<sup>9</sup>, das SGML in den meisten Fällen ersetzt hat, insbesondere bei Anwendungen, die in den letzten Jahren begonnen worden sind. XML verwendet den kompletten Zeichenvorrat von Unicode und ist damit zumindest theoretisch für alle Sprachen und Schriftsysteme gerüstet. Wie bei SGML muss aber die konkrete Anwendung noch spezifiziert werden. Dies geschieht durch eine Document Type Definition (DTD) oder zunehmend durch ein XML-Schema. In beiden Fällen wird festgelegt, welche Elemente in welcher Verschachtelungsform und mit welchen Inhaltstypen vorkommen können. Inzwischen gibt es für XML und die damit zusammenhängenden Standards wie XSLT<sup>10</sup> (eine Skriptsprache für die Umwandlung von XML-Dokumenten) eine große Anzahl von Werkzeugen zu Erstellung und Weiterverarbeitung, so dass keine Hersteller- oder Plattformabhängigkeit zu befürchten ist. XML kann nicht nur, wie im Folgenden weiter beschrieben, zur Auszeichnung kompletter Texte verwendet werden, sondern auch als Codierung für Metadaten-Formate wie RDF oder EAD.

Die wohl bekannteste Anwendung von SGML und XML, nämlich die im WWW verwendete Hypertext Markup Language (HTML), ist für Texteditionen kaum brauchbar, weil ausschließlich auf die Online-Anzeige ausgerichtet. Inzwischen wird XML auch von generellen Office-Programmen als Grundlage des Dokumentenformats verwendet. So beruhen die Formate von OpenOffice.org/StarOffice, iWork von Apple und in der nächsten Programmversion auch Microsoft Office auf XML. Allerdings gewährleistet dies nicht automatisch eine Trennung von Struktur und Form.

Besser berücksichtigt wird diese Forderung bei verschiedenen DTDs,

9 1. Version 1998, 4. Ausgabe 2006, vgl. <<http://www.w3.org/TR/xml/>>.

10 Vgl. <<http://www.w3.org/TR/xslt/>>.

die speziell zur Auszeichnung von Texten entwickelt wurden; beispielhaft sei DocBook genannt.<sup>11</sup> Gezielt mit der wissenschaftlichen Beschreibung und Edition von (bereits vorliegenden) Texten hat sich die Text Encoding Initiative (TEI) beschäftigt, die im Folgenden näher vorgestellt werden soll.

### Text Encoding Initiative

#### *Geschichte*

Die TEI<sup>12</sup> bildete sich 1987, um auf der Grundlage der damals verfügbaren Meta-Auszeichnungssprache SGML ein Auszeichnungssystem für Texte jeder Art zu entwickeln. Ziel war dabei ein gemeinsames Austauschformat, das aber so flexibel sein sollte, dass jede/r Herausgeber/in die Möglichkeit hat, sich aus dem von der TEI bereitgestellten Grundlagen ein für seine/ihre Bedürfnisse passendes Format zusammenzustellen. Nach einigen vorläufigen Versionen lag die erste stabile Fassung der TEI-Richtlinien (Guidelines) mit der Kennung P3 im Jahr 1994 vor, noch auf SGML-Basis. Sie wurde seitdem kontinuierlich weiterentwickelt. Die 2002 veröffentlichte Version P4 beruht erstmals auch auf XML.<sup>13</sup> Aktuell wird an Vorversionen von P5 gearbeitet, das nur noch auf XML ausgerichtet ist und neben einer DTD auch Schemas verwenden kann.<sup>14</sup>

11 Vgl. <<http://www.docbook.org/>>; siehe auch Bunke, Hendrik, *Schreibt strukturiert! XML und Docbook in Sozial- und Geisteswissenschaften* (2005), unter <<http://hbxt.org/edutech/docbook-in-geisteswissenschaften>>.

12 Vgl. <<http://www.tei-c.org/>>. Siehe auch Vanhoutte, Edward, *An introduction to the TEI and the TEI Consortium, 2004*, unter <<http://www.kantl.be/ctb/pub/2004/teillc.htm>>. *Allgemeine Einführungen in deutscher Sprache*: Schmidt, Frieder, *Neuland für die Buchgeschichte – Quellenaufbereitung im Zeitalter des WWW*, in: *Leipziger Jahrbuch für Buchgeschichte* 7 (1997), S. 343-365, besonders S. 351-361; Bader, Winfried, *Was ist die Text Encoding Initiative (TEI)?*, in: Kamzelak (Hg.), *Text-Edition* (wie Anm. 4), S. 9-20; Bruvik, Tone Merete, „Yesterday's information tomorrow“. *Die Text Encoding Initiative (TEI)*, in: *Sichtungen online*, 19.05.2002, unter <<http://purl.org/sichtungen/bruvik-tm-1a.html>>.

13 DeRose, Steven, *XML and the TEI*, in: *Computers and the humanities* 33 (1999), S. 11-30.

14 Zum Stand Oktober 2005: [Burnard, Lou,] *TEI P5 progress report*, unter <<http://www.tei-c.org/Talks/2005/Wuerzburg/p5report.pdf>>.

Die TEI wird durch verschiedene Fachgesellschaften und Forschungseinrichtungen unterstützt, die sich im TEI-Konsortium zusammengefunden haben. Es besteht derzeit (August 2006) aus mehr als 80 Mitgliedern, unter denen erst seit 2006 eines aus dem deutschsprachigen Raum vertreten ist<sup>15</sup>, obwohl auch hier die TEI-Richtlinien in zahlreichen Projekten verwendet werden. TEI beschränkt sich nicht mehr auf die Rolle des Austauschformats, sondern wirkt auch in andere Bereiche der Datenaufbereitung (die W3C-Standards XLink und XPointer gehen auf TEI-Entwicklungen zurück) und hat eine eigene *Research community* hervorgebracht.<sup>16</sup> Die Textauszeichnung nach den TEI-Richtlinien ist insbesondere in den Literaturwissenschaften international weit verbreitet. Aber auch viele für Historiker/innen relevante Materialien sind entsprechend ausgezeichnet.<sup>17</sup>

#### *Prinzipien der TEI-Richtlinien*

Die TEI-Guidelines<sup>18</sup> basieren auf den Erfahrungen bei der Erstellung konventioneller Texteditionen, zwingen also nicht dazu, Formate, die ursprünglich für einen ganz anderen Einsatzzweck entwickelt wurden, mühselig anzupassen. Sie sind modular aufgebaut und können deshalb für verschiedene Textarten adaptiert werden.<sup>19</sup> Neben einem für alle TEI-Dokumente verbindlichen Grundbestand gibt es eigenständige Teile für Prosa, Drama, Lyrik, für kritische Apparate oder analytische Strukturen. Bei zahlreichen Projekten wird eine ursprünglich für didaktische Zwecke

15 Kooperationsprojekt TextGrid, siehe <<http://www.textgrid.de/>>, noch nicht aufgeführt auf der Mitgliederliste unter <<http://www.tei-c.org/membership.php>>.

16 Mylonas, Elli; Renear, Allen, *The Text Encoding Initiative at 10. Not just an interchange format anymore – but a new research community*, in: *Computers and the humanities* 33 (1999), S. 1-9.

17 Die Zahl der TEI-verwendenden Projekte (unter Einschluss derer, die keine vorhandenen Texte, sondern aktuelle Publikationen damit aufarbeiten), liegt weit über jenen 131, die aktuell (28.08.2006) auf der TEI-Website aufgeführt sind, vgl. <<http://www.tei-c.org/Applications/>>.

18 Einführung in die TEI-Richtlinien: Jannidis, Fotis, *TEI in der Praxis*, unter <<http://computerphilologie.uni-muenchen.de/praxis/teipraxis.html>>; Ders., *Wider das Altern elektronischer Texte. Philologische Textauszeichnung mit TEI*, in: *Editio* 11 (1997), S. 152-177.

19 Auch online mit Hilfe des Pizza Chefs für P4, vgl. <<http://www.tei-c.org/pizza.html>>, bzw. von Roma für P5, vgl. <<http://tei.oucs.ox.ac.uk/Roma/>>.

konzipierte Auswahl verwendet, TEI lite, die die am häufigsten verwendeten Elemente zusammenfasst und bei Bedarf auf der Grundlage des vollen Umfangs modifiziert werden kann.<sup>20</sup>

Die TEI-Guidelines beschäftigen sich ausschließlich mit der Struktur des auszuzeichnenden Textes, während im Normalfall keine Festlegungen über Aussehen oder Ausgabeform getroffen werden. Grundprinzip ist es, implizite Informationen explizit zu machen (wie bei konventionell erstellten Texteditionen liegt hier also ein Element der Interpretation vor). Die Richtlinien sehen folgende Bestandteile vor:

- Metadaten (TEI-Header);
- Grundstrukturen des Textes (Kapitel, Strophen, Briefe, Anmerkungen);
- speziell zu formatierende Passagen;
- editorische und semantische Codierung (Anmerkungen des Herausgebers/der Herausgeberin, kritischer Apparat, Namen, sprachliche Besonderheiten usw.);
- durchgehend Spezifizierung durch Attribute;
- möglich ist auch die Integration von Abbildungen und Tabellen sowie Verweise auf weitere Materialien (dazu ist ein eindeutiger Identifier für die Elemente erforderlich).

Voraussetzung für die sinnvolle Anwendung sind eine gewisse Textanalyse und die konsequente Durchführung der Auszeichnung. Wie bei allen Auszeichnungssprachen ist ein systematisches Vorgehen erforderlich. Gerade am Beginn eines Projekts kann dabei der Erfahrungsschatz, der in TEI eingeflossen ist, durchaus hilfreich sein. Eine TEI-Grundstruktur ist relativ schnell zu erzeugen; eine sinnvolle semantische Codierung erfolgt normalerweise mit einem XML-fähigen Editor. Die Auszeichnung ist je nach Bedürfnis und Kapazität verschieden tief möglich; von amerikanischen Bibliotheken entworfene Empfehlungen sehen fünf Ebenen vor, von einer automatischen Konversion bis zur textkritischen Edition.<sup>21</sup>

20 Burnard, Lou; Sperberg-McQueen, C. M., TEI U5. Encoding for interchange. An introduction to the TEI, 1995/2002, unter [http://www.tei-c.org/Lite/teiu5\\_split\\_en.html](http://www.tei-c.org/Lite/teiu5_split_en.html).

21 TEI text encoding in libraries. Guidelines for best encoding practices, Version 2.1,

Die Umwandlung in eine dem gewohnten Lese- und Rezeptionsverhalten entsprechende Form geschieht in der Regel mit Hilfe von Style-Sheet-Sprachen, insbesondere XSLT und XSL-FO<sup>22</sup>; unter Umständen kann auch das eigentlich für HTML entwickelte CSS<sup>23</sup> herangezogen werden. Die Stylesheets sind normalerweise nicht nur für ein Dokument, sondern für eine Vielzahl von Texten zu verwenden. Die Ausgabeformate können dynamisch auf Anfrage (etwa auf einem Webserver) oder im Vorhinein erzeugt werden. Auf diese Weise ist etwa die Herstellung von HTML- oder PDF-Fassungen möglich, aber auch die Generierung von Konkordanzen oder E-Book-Formaten. Die offenen Standards sollten dabei auch für zukünftige Formate offen sein.

Nicht weiter eingehen möchte ich hier auf die Möglichkeit des Zusammenspiels von TEI-codierten Daten und älteren, aber leistungsfähigen Werkzeugen wie TUSTEP<sup>24</sup>, TEX<sup>25</sup> oder Folio Views. Letzteres wird als Retrieval-Oberfläche in zahlreichen elektronischen Editionsprojekten verwendet, steht aber unter den genannten Vorbehalten bezüglich der langfristigen Verfügbarkeit.<sup>26</sup>

27. März 2006, vgl. <http://www.diglib.org/standards/tei.htm>.

22 Vgl. <http://www.w3.org/TR/xsl/>.

23 Vgl. <http://www.w3.org/Style/CSS/>.

24 Vgl. <http://www.uni-tuebingen.de/zdv/tustep/index.html>.

25 Vgl. <http://www.tug.org/>. Ein darauf beruhendes Editionssystem ist etwa CET (Karasch, Bernt, Critical Edition Typesetter (CET). Ein Programmsystem zum Satz textkritischer Editionen auf PCs, in: Kamzelak (Hg.), Text-Edition (wie Anm. 4), S. 87-99), nicht zu verwechseln mit CTE, der ein eigenes Dokumentformat verwendet, aber auch HTML und TEI erzeugen kann (Hagel, Stefan, Zur druckfertigen Edition mit dem eigenen PC – der „Classical Text Editor“, in: Klingenstein u.a. [Hgg.], Umgang [wie Anm. 1], S. 198-201).

26 Die Zuversicht von Monika Seekircher (Der Wittgenstein-Gesamtbriefwechsel in maschinenlesbarer Form, in: Bauer, Werner M.; John, Johannes; Wiesmüller, Wolfgang (Hgg.), „Ich an Dich“. Edition, Rezeption und Kommentierung von Briefen (Innsbrucker Beiträge zur Kulturwissenschaft, Germanistische Reihe 62), Innsbruck 2001, S. 189-204, hier S. 202), ihre mit Folio Views erstellte Wittgenstein-Briefausgabe noch im Jahr 2020 erweitern zu können, erscheint angesichts der wechselvollen Geschichte dieses Programms, das aktuell zur Suchmaschinenfirma FAST gehört (vgl. <http://www.fastsearch.com/l3a.aspx?m=497&amid=3116>), eher zweifelhaft.

### *Anwendungen der TEI-Guidelines*

Die TEI-Richtlinien wurden ursprünglich vor allem im Bereich der Literaturwissenschaft und Linguistik angewendet (insbesondere für große Textcorpora oder Wörterbücher). Ihre Flexibilität macht sie aber auch geeignet für Editionen. Dabei gehen sie über die ursprüngliche Funktion als Austausch- und Sicherungsformat hinaus.

Einige Beispiele aus unterschiedlichen thematischen Bereichen seien genannt, wobei Projekte den Vorzug erhielten, die eine Online-Version anbieten:

- CAMENA (Universität Mannheim/Universität Heidelberg)<sup>27</sup>: neulateinische Texte; HTML auf Grundlage von TEI, Einbindung von Images;
- Jean Pauls Exzerptheft (Universität Würzburg)<sup>28</sup>: diplomatische Transkription;
- Sanger Papers (New York University)<sup>29</sup>: Transformation ›on-the-fly‹ mit PHP/Sablotron<sup>30</sup>;
- Newton Project (Imperial College London)<sup>31</sup>: Benutzer/innen können zwischen diplomatischer und Lese-Transkription wählen;
- eScholarship (University of California)<sup>32</sup>: keine Editionen, sondern aktuelle wissenschaftliche Publikationen auf TEI-Basis;

27 Vgl. <<http://www.uni-mannheim.de/mateo/camena/docs/camena.html>>. Siehe auch Schibel, Wolfgang, CAMENA – Neulateinische Dichtung im World Wide Web, in: Neulateinisches Jahrbuch 3 (2001), S. 211-219; Ders.; Kredel, Heinz, Vom Spezialbestand alter Drucke zum kommentierten Volltextcorpus im WWW. Das Projekt CAMENA (Neulateinische Dichtung Deutschlands), in: Burch, Thomas; Fournier, Johannes; Gärtner, Kurt; Rapp, Andrea (Hgg.), Standards und Methoden der Volltextdigitalisierung. Beiträge des Internationalen Kolloquiums an der Universität Trier, 8./9. Oktober 2001 (Akademie der Wissenschaften und der Literatur, Abhandlungen der Geistes- und Sozialwissenschaftlichen Klasse, Einzelveröffentlichung 9), Stuttgart 2003, S. 187-209.

28 Vgl. <[http://www.uni-wuerzburg.de/germanistik/neu/jp-arbeitsstelle/jpa\\_exz01.htm](http://www.uni-wuerzburg.de/germanistik/neu/jp-arbeitsstelle/jpa_exz01.htm)>.

29 Vgl. <[http://wilde.acs.its.nyu.edu/sanger\\_dev/documents/search.php](http://wilde.acs.its.nyu.edu/sanger_dev/documents/search.php)>.

30 Zimmerman, Matthew, Publishing XML files on the web (2003), vgl. <[http://www.nyu.edu/its/pubs/connect/fall03/zimmerman\\_xml.html](http://www.nyu.edu/its/pubs/connect/fall03/zimmerman_xml.html)>.

31 Vgl. <<http://www.newtonproject.ic.ac.uk/>>. Zur Technik: Dunning, Alastair, The Newton Project. Implementing and exploiting XML, 7. November 2005 <[http://ahds.ac.uk/\\_print\\_/creating/case-studies/newton/index.htm](http://ahds.ac.uk/_print_/creating/case-studies/newton/index.htm)>.

32 Vgl. <<http://www.cdlib.org/programs/escholarship.html>>.

- Sichtungen (Österreichisches Literaturarchiv der Österreichischen Nationalbibliothek)<sup>33</sup>: elektronische Zeitschrift mit TEI-Master, HTML- und Druckfassung generiert durch XSL-Stylesheets.

Nur am Rande sei darauf hingewiesen, dass auch das umfangreiche, aber wissenschaftlichen Kriterien bisher nicht genügende Textarchiv des Project Gutenberg<sup>34</sup> seit kurzem TEI-codierte Texte aufnimmt.

### **Erfahrungsbericht: Computereditionen der Bibliothek für Bildungsgeschichtliche Forschung**

Die Bibliothek für Bildungsgeschichtliche Forschung (BBF)<sup>35</sup> ist die größte pädagogische Spezialbibliothek Deutschlands und gehört zum Deutschen Institut für Internationale Pädagogische Forschung, einer Einrichtung der Leibniz-Gemeinschaft. Als Forschungsbibliothek hat sie außer den eigentlichen bibliothekarischen und dokumentarischen Aufgaben einen Schwerpunkt bei der Zusammenarbeit mit der Wissenschaft in Form von Forschungs- und Publikationsprojekten, Tagungen und Ausstellungen. Neben Digitalisierungs- und Fachinformationsprojekten ist die BBF seit 2001 auf dem Gebiet der Computereditionen aktiv.

#### *Werkausgabe Adolf Reichwein*

Die Werkausgabe des Pädagogen und Widerstandskämpfers Adolf Reichwein (1898–1944) war ursprünglich als konventionelle Edition geplant, die in fünf Bänden alle pädagogischen Schriften umfassen sollte.<sup>36</sup> Bei Beginn der von der Deutschen Forschungsgemeinschaft (DFG) und weiteren Sponsoren geförderten Arbeiten wurde jedoch eine Erfassung aller, auch der nicht-pädagogischen Schriften Reichweins ins Auge gefasst, die in Form einer CD-ROM den Bänden der Werkausgabe beigegeben werden sollen. Von vornherein war für die Erfassung aller Texte die Verwendung von TEI vorgesehen, das somit auch als Grundlage der weiterhin vorgesehenen Druckausgabe dienen soll. Das wissenschaftliche und technische Editions-

33 Vgl. <<http://purl.org/sichtungen/>>.

34 Vgl. <<http://www.gutenberg.org/>>.

35 Vgl. <<http://www.bbf.dipf.de/>>.

36 Wissenschaftliche Bearbeiter: Ullrich Amlung, Karl Christoph Lingelbach. Die Edition wird vom Adolf-Reichwein-Verein unterstützt.

konzept wurde im Februar 2003 auf einem Expertenworkshop in der BBF vorgestellt, der zu einigen Modifikationen führte. Insbesondere der Umfang der Kommentierung wurde reduziert, um den durch die DFG-Finanzierung gesetzten Zeitrahmen nicht zu überschreiten. Die ersten beiden Bände sollen Ende 2006 in druckfertiger Form vorliegen. Noch nicht abschließend geklärt ist, ob Teile oder das gesamte Textcorpus auch über das Internetangebot der BBF zur Verfügung gestellt werden können.

Die Werke Reichweins sind zum großen Teil nur noch in der publizierten Fassung erhalten, selten als Schreibmaschinentypskript; Handschriftliches gibt es praktisch überhaupt nicht (Reichwein wurde während des Krieges ausgebombt). Die Texterfassung erfolgte teilweise über OCR, bei schlechten oder frakturgedruckten Vorlagen über manuelle Erfassung. Die daraus resultierenden Dokumente im DOC-Format von Microsoft Word wurden, nachdem Möglichkeiten der automatischen Konvertierung getestet, aber verworfen wurden, als unformatierter Text in den XML-Editor XMetaL geladen und dort mit einer zunächst elementaren Auszeichnung nach TEI lite versehen. Neben den eigentlichen Textabschnitten wurden dabei auch Fußnoten und Tabellen berücksichtigt. Diese Phase, die auch mit Verwendung von studentischen Hilfskräften erfolgte, ist 2004 abgeschlossen worden, so dass seitdem alle Texte Reichweins zumindest in einer minimalen TEI-Struktur vorhanden sind. XMetaL hat sich, nicht zuletzt durch seine wahlweise an eine normale Textverarbeitung erinnernde Benutzeroberfläche, in diesem und den im Folgenden vorgestellten Projekten bewährt, auch wenn die Wahl eines XML-Editors immer auch eine Frage der persönlichen Vorliebe ist.<sup>37</sup>

Eine inhaltliche Auszeichnung wurde bisher bei einem Teil der Texte vorgenommen. Dabei sind nach Entscheidung der wissenschaftlichen Bearbeiter signifikante Personen- und Ortsnamen, gelegentlich auch Institutionen, durch die entsprechenden TEI-Elemente gekennzeichnet worden. Vorgesehen ist die automatische Generierung eines entsprechenden Index

<sup>37</sup> Vgl. <http://www.xmetal.com/>. Wer das Arbeiten in einer code-orientierten Ansicht nicht scheut, findet etwa bei oXygen unter <http://www.oxygenxml.com/> eine gute Unterstützung der TEI und aktueller Technologien wie XSLT und XML Schema.

für die Druckausgabe, der verbunden ist mit einem Glossar der Personennamen, das Teile des ursprünglich geplanten Kommentars ersetzt.

### Briefausgabe Friedrich Fröbel

Die bisher nur auszugsweise edierten Briefe des Pädagogen Friedrich Fröbel (1782–1852) wurden seit 1991 mit Unterstützung der DFG durch Helmut Heiland (Universität Duisburg) erfasst und transkribiert. Die mit einem knappen textkritischen Apparat, aber keinem Kommentar versehene Transkription erfolgte im DOC-Format, aufgrund der langen Bearbeitungszeit in verschiedenen Versionen.

Das Corpus der Fröbel-Briefe besteht aus 1.661 bekannten Briefen oder Brieffragmenten, die im Word-Format etwa 7.300 Seiten umfassen.

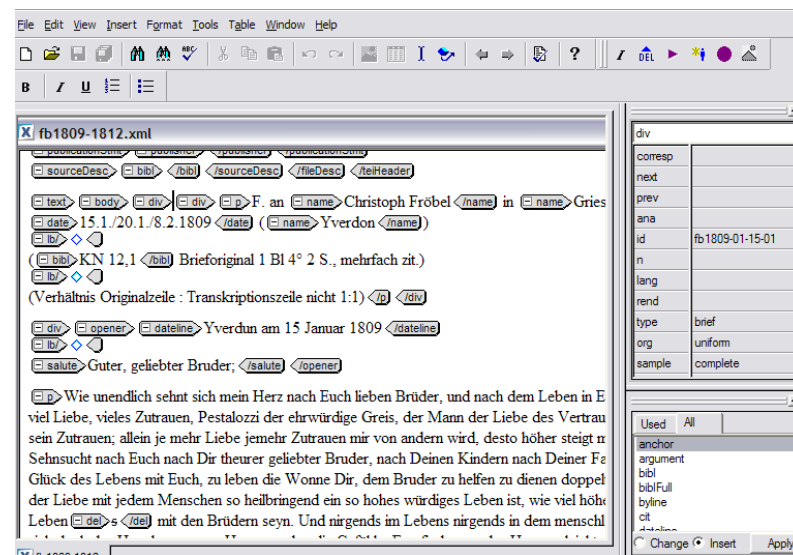


Abbildung 1: Text eines Fröbelbriefes im Editor XMetaL

Angesichts dieses Umfangs war eine komplette Veröffentlichung in Druckform nicht praktikabel, so dass eine Computeredition ins Auge gefasst wurde, bei der die BBF ihre bei der Reichwein-Ausgabe gewonnenen

technischen Erfahrungen einbringt. Seit September 2005 werden in einem zweijährigen, von der DFG geförderten Projekt die in etwa 50 Word-Dateien vorliegenden Texte in TEI lite konvertiert. Dabei wird als erster Schritt eine automatische Konversion mit Hilfe des TEIOO-Filters von Sebastian Rahtz (Universität Oxford) durchgeführt.<sup>38</sup> Da die Dateien im Word-Format aufgrund der langen Bearbeitungszeit aber zahlreiche Formatierungsunregelmäßigkeiten aufweisen und rein visuell orientiert waren, müssen umfangreiche Nacharbeiten erfolgen. Dabei werden die einzelnen Briefe voneinander abgetrennt, innerhalb der Briefe editorische Einleitung und Transkription, bei dieser wiederum Einleitungs- und Schlussformeln gekennzeichnet.

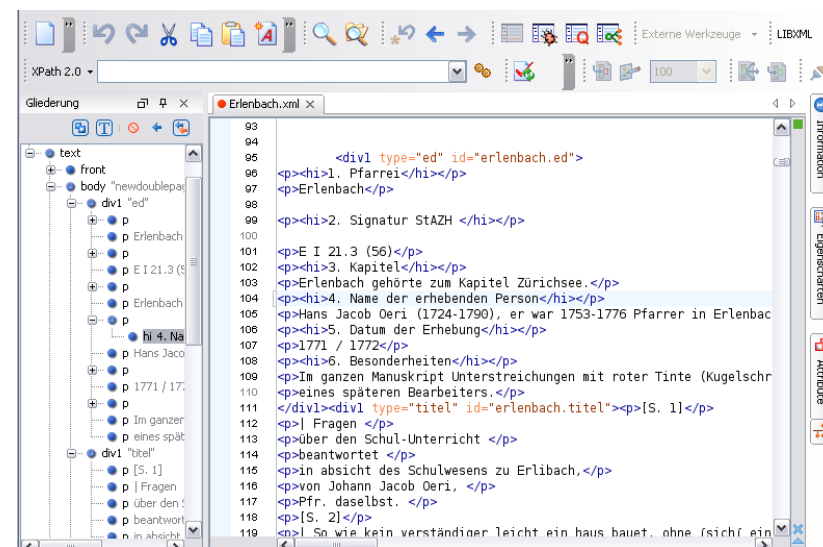
Die Transkription hatte ursprünglich versucht, das Erscheinungsbild der handschriftlichen Vorlagen zumindest näherungsweise wiederzugeben. Daher finden sich zahlreiche Einzüge, horizontale oder vertikale Leerräume und Schriftformatierungen wie hochgestellt, unterstrichen (einfach, doppelt und vereinzelt sogar dreifach) oder kursiv (als Kennzeichen des Wechsels zwischen deutscher und lateinischer Handschrift). Bei der Umsetzung in TEI-Stukturen ist dabei oftmals die Frage aufgetaucht, wie sich diese Formatierungseigenschaften in logische Strukturen umsetzen lassen. Eine 1:1-Identität des Endergebnisses mit der Word-Fassung (die wiederum ja schon eine Abstraktion der Handschrift darstellt) wird sich nicht erreichen lassen und ist auch nicht angestrebt. Der textkritische Apparat, der Streichungen, Einfügungen und unsichere Lesungen betrifft, wird weitgehend in die entsprechenden TEI-Elemente umgesetzt. Hier hat sich auch die Notwendigkeit gezeigt, möglichst früh während der Bearbeitung in Absprache mit den wissenschaftlichen Bearbeitern/innen klare Entscheidungen für Zweifelsfälle zu treffen. Eine gewisse Vertrautheit mit den Eigenheiten der handschriftlichen Vorlage ist dabei sinnvoll, auch wenn die eigentliche Transkription bereits erfolgt ist. Aktuell ist noch die Frage offen, ob weiterhin mit TEI lite gearbeitet wird oder die DTD auf der Grundlage des Gesamtumfangs der TEI modifiziert werden soll. Die Veröffentlichung ist für 2007 über das Internetangebot der BBF vorgese-

38 Vgl. <<http://www.tei-c.org/Software/teioo/>>.

hen. Als Ergänzung ist ein Einführungsband vorgesehen, der auch ausgewählte Briefe abdrucken wird, aber kein eigentlicher Bestandteil der Computeredition ist.

#### Weitere Projekte und Planungen

Bei einem Editionsprojekt des Instituts für Historische Bildungsforschung Pestalozzianum Zürich<sup>39</sup> gibt die BBF im Rahmen einer schon länger bestehenden Kooperation technische Unterstützung. Die Transkription einer 1771/1772 durchgeführten Fragebogenaktion bei den Pfarrern des Kantons Zürich über den Zustand ihrer Landschulen wird, ähnlich wie bei den Fröbelbriefen, aus einer vorliegenden Word-Fassung in TEI lite umgesetzt. Die Veröffentlichung des gesamten Textes erfolgt 2006 als CD-ROM, die einem Kommentarband beigelegt ist.



Die TEI-Auszeichnung und elektronische Veröffentlichung weiterer

39 Vgl. <<http://ihbf.phzh.ch/>>.

bildungshistorischer Quellenbestände aus dem 19. und 20. Jahrhundert ist geplant, nachdem mehrere Wissenschaftler/innen Interesse an einer Kooperation geäußert haben. In einem Pilotprojekt wird derzeit ferner ein Mitteilungsblatt nach den TEI-Richtlinien ausgezeichnet, um die Möglichkeit der langfristigen Archivierung auch aktueller Publikationen nach diesem Modell zu testen.

#### *Erste Ergebnisse und Erfahrungen*

Die Erfahrungen der BBF zeigen, dass es möglich ist, sich rasch in die Eigenheiten der Textauszeichnung nach TEI einzuarbeiten. Bei den bisherigen Projekten gab es immer nur eine/n Hauptbearbeiter/in, so dass kaum Probleme bei der Arbeitsorganisation und Versionskontrolle anfielen. Zukünftig wird erwartet, dass mehrere Personen gleichzeitig oder nacheinander an Texten arbeiten; damit ergibt sich die Notwendigkeit für ausführliche und eindeutige schriftliche Auszeichnungsanweisungen, quasi „Anwendungsbestimmungen“ für die TEI-Guidelines.

Die in der Community schon lange diskutierten Mängel des SGML/XML-basierten TEI-Markups werden auch bei den Editionsprojekten der BBF deutlich: Logische Strukturen lassen sich oft nicht von physischen trennen<sup>40</sup>; auch die Anordnung der Schrift und weiterer grafischer Elemente auf dem Papier kann Bedeutung tragen. Die TEI-Richtlinien waren bisher weitgehend auf die inhaltlichen Strukturen ausgerichtet, enthalten im neuesten Entwurf P5 aber auch einen Abschnitt zur „Manuscript Description“.<sup>41</sup> Überlappende Strukturen (etwa Sinnabschnitte und Manuskript- oder Druckseiten) lassen sich auf XML-Grundlage nicht gleichermaßen gut auszeichnen; es gibt aber einige Versuche, dieser

40 Siehe dazu auch: Vanhoutte, Edward, Display or argument? Markup and visualization for electronic scholarly editions, in: Burch u.a. (Hgg.), Standards (wie Anm. 27), S. 71-96.

41 Die andauernde Weiterentwicklung der TEI-Richtlinien und der zugrunde liegenden Standards können natürlich auch Zweifel an der Dauerhaftigkeit einer darauf aufbauenden Edition erzeugen: Rischer, Tobias, Eine säurefreie elektronische Edition des Ulysses. Bestandsaufnahme und einige Vorschläge, in: Henkes, Christiane u.a. (Hgg.), Schrift – Text – Edition. Hans Walter Gabler zum 65. Geburtstag (Beihefte zu Editio 19), Tübingen 2003, S. 339-348.

bekannten Beschränkung zu begegnen.<sup>42</sup> Eine mögliche Abhilfe für die genannten Mängel wäre die Integration von Images (elektronischen Faksimiles), wie bei CAMENA und vielen anderen Projekten bereits erfolgt.<sup>43</sup>

#### **Ausblick: Zusammenarbeit von Wissenschaft und Bibliotheken in der elektronischen Quellenedition**

Viele Editions- und Textauszeichnungsprojekte leiden unter mangelhafter technischer Unterstützung. Es ist aber auch kaum vorstellbar, jedes Projekt mit der entsprechenden Ressourcenausstattung, insbesondere der personellen, zu versehen. Hier ist die Unterstützung durch Kompetenzzentren denkbar, die sinnvollerweise in einem klar definierten disziplinären Kontext agieren. Sie können dadurch unter Umständen näher an den konkreten Bedürfnissen der Wissenschaftler/innen sein als allgemeine „Electronic Text Center“<sup>44</sup>, auch wenn etwa das Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften an der Universität Trier erfolgreich Projekte aus verschiedenen Fachgebieten betreut.<sup>45</sup>

Eine wichtige Rolle kann hier den Forschungsbibliotheken zukommen, die sich durch die enge Verbindung mit der Wissenschaft definieren.<sup>46</sup> (Wenn im Folgenden die Bibliotheken im Vordergrund stehen, ist das dem persönlichen Hintergrund geschuldet. Natürlich können auch Computerzentren, Archive oder spezielle Einrichtungen ähnliche Dienstleistungen erbringen.) Durch diese Zusammenarbeit mit den Erzeugern elektronischer Angebote, hier also besonders von Editionen, befinden sie sich in einer anderen Situation als etwa regional definierte Pflichtexemp-

42 Ein Vorschlag, die materiellen Eigenschaften alter Drucke mit zu erfassen: Bauman, Syd; Catapano, Terry, TEI and the encoding of the physical structure of books, in: Computers and the humanities 33 (1999), S. 113-127.

43 Siehe auch Flanders, Julia, Trusting the electronic edition, in: Computers and the humanities 31 (1998), S. 301-310.

44 Seaman, David, The Electronic Text Center at the University of Virginia Library, in: Burch u.a. (Hgg.), Standards (wie Anm. 27), S. 97-107.

45 Vgl. <<http://germazope.uni-trier.de/Projects/KoZe2>>.

46 Grundlegend: Knoche, Michael, Die Forschungsbibliothek. Umriss eines in Deutschland neuen Bibliothekstyps, in: Bibliothek. Forschung und Praxis 17 (1993), S. 291-300.

larbibliotheken, die mit heterogenen digitalen Publikationen rechnen müssen, auf deren Gestaltung sie nur beschränkt, wenn überhaupt, Einfluss haben. Neben einer disziplinären Zuordnung ist natürlich auch eine Spezialisierung auf bestimmte Epochen oder Materialgattungen denkbar.<sup>47</sup>

Bibliotheken können ihre Dienste insbesondere auch bei der Bereitstellung der Publikationen und der Einbindung in umfassendere Angebote erbringen; hier kommen ferner Themen wie Metadaten und eindeutige Identifizier ins Spiel, die ebenfalls in den Kompetenzbereich von Bibliotheken und Archiven fallen, in diesem Abriss aber nicht behandelt werden konnten. Gleiches gilt schon bei der Textauszeichnung für die Regularisierung von Namensformen, wo die Verwendung von Normdaten sehr hilfreich sein kann.<sup>48</sup> Besonders wichtig ist natürlich der Bereich der Langzeitarchivierung: Bei selbst betreuten Veröffentlichungen ist eine datenträgerunabhängige Migration anders als bei vielen Fremdprodukten sinnvoll durchführbar, weil dieser Aspekt von vornherein berücksichtigt werden kann. Das Problem proprietärer CD-ROMs, das die Archivbibliotheken bereits erreicht hat (von den Disketten früherer Jahre zu schweigen) ist damit zwar nicht beseitigt, aber zumindest ein möglicher Lösungsweg aufgezeigt, der idealerweise Vorbildcharakter haben könnte.

Das aktuelle Schlagwort des Open Access, also des freien Zugangs zu wissenschaftlichen Publikationen, ist im Bereich der Quelleneditionen, die bestenfalls einen marginalen kommerziellen Wert haben, natürlich äußerst relevant. Neben der besseren Sichtbarkeit im Internet, insbesondere in allgemeinen Suchmaschinen, erlauben offene Formate und entsprechende Lizenzen auch die Spiegelung auf weiteren Servern, was unter Umständen ebenfalls ein Element einer langfristigen Archivierungsstrategie sein

47 Für Altbestandsbibliotheken siehe Stäcker, Thomas, XML für alte Drucke, in: Benkert, Hannelore; Rosenberger, Burkard; Dittrich, Wolfgang (Hgg.), 92. Deutscher Bibliothekartag in Augsburg 2002. Die Bibliothek zwischen Autor und Leser (Zeitschrift für Bibliothekswesen und Bibliographie, Sonderheft 84), Frankfurt am Main 2003, S. 259-275.

48 Diesen Punkt hebt etwa hervor Sukovic, Suzana, Beyond the scriptorium. The role of the library in text encoding, in: D-Lib magazine 8 (2002), Nr. 1, unter <<http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/january02/sukovic/01sukovic.html>>.

könnte. Unbedingt erforderlich ist es dafür, neben einer aufbereiteten (HTML- oder PDF-) Fassung auch den TEI-codierten XML-Text bereitzustellen. Die bisher noch oft übliche Veröffentlichung auf einer (separaten oder meist zusammen mit einem Buch angebotenen) CD-ROM gewährleistet zwar eine stabile Version, ist aber unter langfristiger Betrachtungsweise nicht weniger problematisch, als es eine potenziell verschwundene Website wäre.

In der Praxis wird man natürlich Rücksicht nehmen müssen auf die gewohnten Rezeptionsgewohnheiten, Vorbehalte wegen der Manipulierbarkeit digitaler Publikationen und traditionelle Verlagsbeziehungen, die man aber durchaus auch für Open Access und die Verwendung standardisierter Formate ausnutzen kann. Die Hybridedition in gedruckter und elektronischer Form hat sich in der wohl noch lange andauernden Übergangszeit, in der wir uns befinden, durchaus bewährt, und offenbar können sich viele Wissenschaftler/innen und Bibliothekar/innen allmählich mit dem Gedanken anfreunden, dass eine auf lange Dauer angelegte Veröffentlichung wie eine Edition nicht notwendigerweise ein gedrucktes Buch sein muss.

\*\*\*

*Dr. Stefan Cramme ist stellvertretender Bibliotheksleiter der Bibliothek für Bildungsgeschichtliche Forschung des Deutschen Instituts für Internationale Pädagogische Forschung; Arbeitsschwerpunkte sind derzeit: Computereditionen, Fachinformation für die historische Bildungsforschung. E-Mail: [cramme@bbf.dipf.de](mailto:cramme@bbf.dipf.de)*

## VERTEILTE DIGITALE INKUNABELBIBLIOTHEK – EIN BAUSTEIN ZUR GESAMTDIGITALISIERUNG ALLER INKUNABELAUSGABEN

von Timo Steyer

*Auf der Internetseite des Forschungsprojektes „Verteilte Digitale Inkunabelbibliothek“ (vdIb) werden bis Mitte des Jahres 2006 über 1.200 digitalisierte und mit Metadaten erschlossene Inkunabeln der Forschung zur Verfügung gestellt sein.<sup>1</sup> Der Artikel berichtet über die Projektarbeit und die Möglichkeiten, welche die Projektseite für die Benutzung der Digitalisate offeriert. Anhand dieser Ausführungen sollen ferner die Grenzen, welche eine quantitative Digitalisierung bietet, diskutiert werden. Dabei wird auch der Frage nachgegangen, inwieweit das Projekt als Muster für eine zukünftige Gesamtdigitalisierung aller Inkunabelausgaben dienen kann.*

\*\*\*

### Einleitung

Die retrospektive Digitalisierung historischer Quellen erfährt in jüngster Zeit eine Konjunktur, die nicht nur in der Förderung durch die Deutsche Forschungsgemeinschaft (DFG) zum Ausdruck kommt, sondern auch in der Vielzahl von internationalen Digitalisierungsprojekten sichtbar wird.<sup>2</sup> Aufgrund des mittlerweile erreichten technischen Stands können diese Vorhaben quantitative und qualitative Ergebnisse erzielen, die vor wenigen Jahren noch als wesentlich längerfristige Zukunftsperspektiven galten. Ein

<sup>1</sup> Projekt-Website: <<http://inkunabeln.ub.uni-koeln.de/>>.

<sup>2</sup> Vgl. dazu die Linksammlung der Zentral- und Landesbibliothek Berlin zu Digitalisierungsprojekten: <<http://linksammlungen.zlb.de/1.2.2.61.0.html>> (14.02.2006).

Ende dieser Entwicklung ist bei weitem noch nicht in Sicht. Doch diese Dynamik wirft auch Fragen auf: Wie korrespondiert die Möglichkeit der quantitativen Digitalisierung mit der Herstellung einer möglichst großen Benutzerfreundlichkeit und inhaltlicher Erschließung der digitalisierten Quellen? Und wie können bereits bestehende Internetressourcen in neue Projekte eingebunden werden, um Synergieeffekte zu erzielen? Die folgenden Ausführungen sollen anhand des Forschungsprojektes der „Verteilten Digitalen Inkunabelbibliothek“ (vdIb) diesen Fragen nachgehen und daher zum einen aufzeigen, wie das Projekt mit anderen Ressourcen verknüpft ist. Zum anderen soll durch die Darstellung der inhaltlichen und technischen Vorgehensweisen bei der Projektarbeit ein Eindruck vermittelt werden, wie eine Digitalisierung von Inkunabeln unter einem quantitativen Fokus vonstatten gehen kann. Dabei soll das Projekt als Diskussionsgrundlage, aber auch als Vorschlag für ähnliche Digitalisierungsprojekte dienen.

### Gegenstand und Philosophie der vdIb

Die vdIb ist ein von der DFG gefördertes Kooperationsprojekt zwischen der Universität Köln, der Universitäts- und Stadtbibliothek Köln sowie der Herzog August Bibliothek (HAB) in Wolfenbüttel, welches eine möglichst repräsentative Digitalisierung der jeweiligen Inkunabelbestände unter Berücksichtigung konservatorischer Gutachten anstrebt.

In Wolfenbüttel wurden im Verlauf des Projektes 667 Inkunabeln mit insgesamt 123.241 Seiten digitalisiert. Dabei entspricht die Gesamtmenge der digitalisierten Inkunabeln etwa fünf bis sechs Prozent der weltweit noch existierenden Wiegendrucke. Diese Zahlen machen deutlich, dass der Schwerpunkt des Projektes auf der Erreichung quantitativer Zielvorgaben lag. Die Idee hierbei war, dass es der Forschung am ehesten nutzt, wenn eine möglichst hohe Anzahl von benutzbaren Quellen im Internet verfügbar ist, auch wenn dabei Beschränkungen in der inhaltlichen Erschließung in Kauf genommen werden müssen.

Zur Produktionsbeschleunigung der Digitalisierung wurde der Inkunabelbestand zwischen den beiden Projektpartnern entsprechend dem Druckjahr aufgeteilt. So wurden in Köln Inkunabeln ausgewählt, welche

vor dem Jahr 1485 gedruckt worden sind. In Wolfenbüttel dagegen wurden Inkunabeln aus der Zeit von 1485 bis 1500 digitalisiert. Die Auswahl des Jahres 1485 als Trennpunkt ergab sich aufgrund der zeitlichen Verteilung der Inkunabelbestände beider Sammlungen, die sich so zu fast gleich großen Stückzahlen auf die Projektteilnehmer verteilten. Der Schwerpunkt lag bei beiden Sammlungen auf Inkunabeln aus deutschen Druckorten. Freilich sind trotzdem zahlreiche Inkunabeln aus dem italienischen, niederländischen und französischen Raum vertreten. Thematisch vermitteln die Digitalisate in ihrer Gesamtheit aufgrund der zur Verfügung stehenden umfangreichen Sammlungen einen guten Einblick in die vielfältige Thematik der Inkunabelzeit. Neben der Digitalisierung bildete die Bereitstellung einer möglichst benutzerfreundlichen Handhabung der Inkunabeln einen zweiten Schwerpunkt im Rahmen des Projektes. Dieser Anspruch wurde nicht nur an die Projektseite im Internet gestellt, sondern auch auf die Digitalisate selber bezogen. Aus diesem Grund wurden Teile der Texte transkribiert und drucktypische sowie spezifische Eigenschaften der einzelnen Inkunabeln vermerkt (siehe dazu das Kapitel „Inhaltliche Bearbeitung der Inkunabeln“). Des Weiteren ist die Benutzung der bereitgestellten Inkunabeln maßgeblich von der Qualität der Digitalisierung abhängig, weshalb im Folgenden ausführlich auf die technischen Rahmenbedingungen der Digitalisierung eingegangen werden soll.

### Technische Rahmenbedingungen

Die Digitalisierung der Inkunabeln inklusive der Einbände erfolgte nach den DFG- Praxisregeln mit mindestens 300 dpi.TIFF. Dabei sind die Digitalisate im JPEG-Format in den drei Auflösungsmodi 600, 1024 und 2000 Pixel verfügbar. Insgesamt werden allein für die Rohdatenmenge des Gesamtprojektes ungefähr 3,5 Terrabyte (TB) Speicherplatz benötigt, für die Sicherstellung der Daten noch einmal die gleiche Speichermenge. Die Digitalisate wurden in Form von 24 BitDateien langfristig gespeichert. Das langfristige Speicherkonzept dient nicht nur der Datensicherung, sondern durch die Speicherung der Rohdaten kann auf zukünftige technologische Fortschritte bei der Hardwareentwicklung Rücksicht genommen werden. Das folgende Beispiel aus einer französischen Ausgabe der Genealogiae

deorum von Boccaccio<sup>3</sup> soll die Qualität der Digitalisate verdeutlichen:

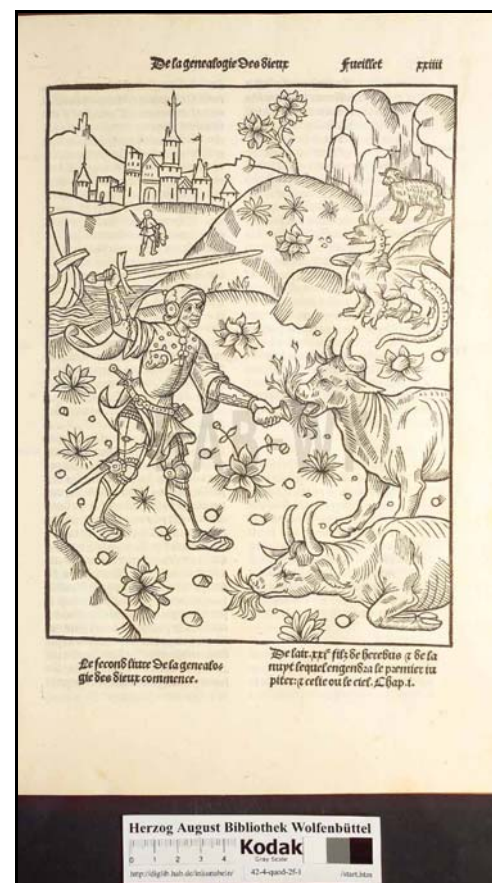


Abbildung 1: Digitalisat aus Giovanni Boccaccios *Boccace de la geneologie Des Dieux*, 1498

<sup>3</sup> Giovanni Boccaccio, *Boccace de la geneologie Des Dieux*, Paris 1498, hier S. 49 der digitalen Ausgabe: <<http://diglib.hab.de/wdb.php?dir=inkunabeln/42-4-quod-2f-1>> (14.02.2006).

Voraussetzung für eine preiswerte und schnelle Digitalisierung des empfindlichen Schriftguts war eine spezielle Buchwippe, der so genannte Wolfenbüttler Buchspiegel. Diese Buchwippe stellt einen wesentlichen Beschleunigungsfaktor dar, weil sie zügige Aufnahmen mit einem 45-Grad-Öffnungswinkel erlaubt. Dieses Verfahren stellt eine schonende Bearbeitung des Materials sicher; bei herkömmlichen Methoden muss das Werk häufig in einem 180° Winkel geöffnet werden, was bei empfindlichen Inkunabelausgaben zu Beschädigungen des Einbandes geführt hätte. Das Material wurde farbig digitalisiert, da die Verwendung von farbigen Partien in den Inkunabeln bewusst zur Textstrukturierung eingesetzt wurde und diese Informationen so für die Forschung zugänglich sind. Auch für die buchhistorische Forschung sollte die Benutzung der Digitalisate einen Gewinn darstellen, da das Material in der angegebenen Qualität die normale Lektüre, auch bei kleinbuchstabigem Druck, ermöglicht. Gleichfalls sollten die Digitalisate in dieser Beschaffenheit für typenkundliche Untersuchungen ausreichen, solange keine speziellen visuellen Behandlungsschritte vorausgesetzt werden wie zum Beispiel Durchlichtaufnahmen zur Wasserzeichenbestimmung.

#### **Inhaltliche Bearbeitung der Inkunabeln: Vom Rohdatensatz zum erschlossenen Digitalisat**

Bereits eingangs wurde das Anliegen der vdlb erwähnt, die bereitgestellten Digitalisate möglichst anwendungsfreundlich im Netz zur Verfügung zu stellen. Aus diesem Grund wurden die fertigen Digitalisate zunächst unter Verwendung eines Image-Viewers mit dem Namen TOC (Table of Content)-Editor<sup>4</sup> durchgesehen, um potenzielle Fehler bei der Digitalisierung festzustellen. Die Struktur der Dateien der Digitalisate nach dieser Bearbeitungsphase soll anhand der Seiten 15 bis 20 einer Ausgabe der Epigramma des Ausonius<sup>5</sup> vorgestellt werden:

4 Eine Dokumentation zu den Einsatzmöglichkeiten und dem Aufbau des TOC-Editors befindet sich auf der Projektseite: <<http://inkunabeln.uni-koeln.de/>> unter den Punkten Projekt vdlb/ Erschließung/TOC-Editor.

5 Ausonius, Avsonii Peonii Poetae Disertissimi Epigrammata, Venedig 1496.

```
[...]
    <div level="0" >
    <page>
    <image extref="00015.jpg" seqno="019"
nativeno="15" />
    </page>
    <page>
    <image extref="00016.jpg" seqno="020"
nativeno="16" />
    </page>
    <page>
    <image extref="00017.jpg" seqno="021"
nativeno="17" />
    </page>
    <page>
    <image extref="00018.jpg" seqno="022"
nativeno="18" />
    </page>
    <page>
    <image extref="00019.jpg" seqno="023"
nativeno="19" />
    </page>
    <page>
    <image extref="00020.jpg" seqno="024"
nativeno="20" />
    </page>
[...]
```

Wie der Name TOC-Editor schon nahe legt, wurden die Digitalisate mit diesem Programm nicht nur auf mögliche Digitalisierungsfehler untersucht, sondern auch weiter bearbeitet. Das Programm erlaubt ein gewisses Maß an inhaltlicher Erschließung der digitalisierten Inkunabeln, das allerdings vom Erschließungsausmaß nicht mit Handschriftenbeschreibungen verglichen werden sollte. Eine solch detaillierte Bearbeitung der Digitalisate hätte in keinem Verhältnis zu den Projektzielen und -vorgaben gestanden. Vielmehr ging es um die Bereitstellung von Navigationshilfen, die es den Benutzern/innen ermöglichen, sich schnell in die Textstruktur einzufinden und so den gewünschten Textabschnitt in kurzer Zeit zu finden.

In der Regel wurde daher mindestens alle zehn Seiten ein Textab-

schnitt des Digitalisates transkribiert. Diese Regel wurde jedoch nicht stringent beachtet: Wo texteigene Übersichtsstrukturen vorhanden waren, wurde auf diese zurückgegriffen, um so eine möglichst große Nähe zur Quellenstruktur und deren Inhalt zu wahren. Doch gab es vor allem bei den frühen Inkunabeln häufig Ausgaben, die weder Inhaltsübersichten noch markante Textabschnitte in verwendbaren Seitenabständen aufwiesen. In diesen Fällen erfolgte die Transkription der betreffenden Seitenanfänge. Maximal wurden drei Texteinträge pro Seite aufgeführt, also zum Beispiel die Überschrift des anfangenden Buchteils, dann die Überschrift des Kapitels und dessen erster Satz. Zusätzlich erlaubte es das Programm, bestimmte Kommentare und Schlagworte zu den einzelnen Seiten zu verfassen. Auf diese Weise konnten fehlende oder vertauschte Lagen, Druckfehler, das Vorhandensein eines Inhaltsverzeichnisses usw. festgehalten werden. In der gleichen Weise wurde bei denjenigen Inkunabeln, für die eine moderne Edition greifbar war, die Kommentierfunktion verwendet, um die transkribierten Passagen in der jeweiligen Edition nachweisen zu können. Bei einigen Literaturarten wurden auch themenspezifische Merkmale der dazugehörigen Inkunabeln mittels der Kommentier- bzw. Schlagwortfunktion aufgeführt. So wurden zum Beispiel bei der Sermonesliteratur in entsprechenden Abständen die Feier- und Heiligentage in der gebräuchlichen deutschen Bezeichnung wiedergegeben und zusätzlich die erste vorgeschlagene Bibelstelle für die Predigt des jeweiligen Tages entsprechend den Loccumer Richtlinien aufgeführt.

Außer den eigenen Eintragungen erlaubte es das Programm, vorhandene Eigenheiten der jeweiligen Inkunabel – wie Provenienzen, Kolophone, Musiknoten, Illustrationen, Schmuckinitialien, Druckerzeichen, Annotationen und das Titelblatt – festzuhalten. Diese Eigenschaften wurden allerdings – aufgrund der quantitativen Ausrichtung der vdIb – lediglich vermerkt und nicht ausgewertet. So erfahren die Nutzer/innen zwar, auf welchen Seiten sie zum Beispiel Annotationen finden, aber deren Inhalt und Gegenstand müssen sie selbst ergründen. Die aufgezählten Erschließungsfunktionen helfen ihnen bei der schnellen Orientierung innerhalb des Textes. Die eingegebenen Transkriptionen und die vermerkten Texteingenschaften können entweder mit den entsprechenden Seiten

zusammen oder aber ohne diese in einer Übersicht aufgerufen werden. Die Digitalisate bieten somit den Benutzern/innen teilweise eine bessere Möglichkeit mit dem Inhalt des Textes zu arbeiten als die Originale. Als Beleg dafür sei der bereits vorgestellte Abschnitt der Ausonius-Ausgabe in seiner fertigen Bearbeitung angeführt, vorangestellt ist die 15. Seite der digitalen Ausonius-Ausgabe<sup>6</sup> mit Bearbeitungsfenster des TOC-Editors:

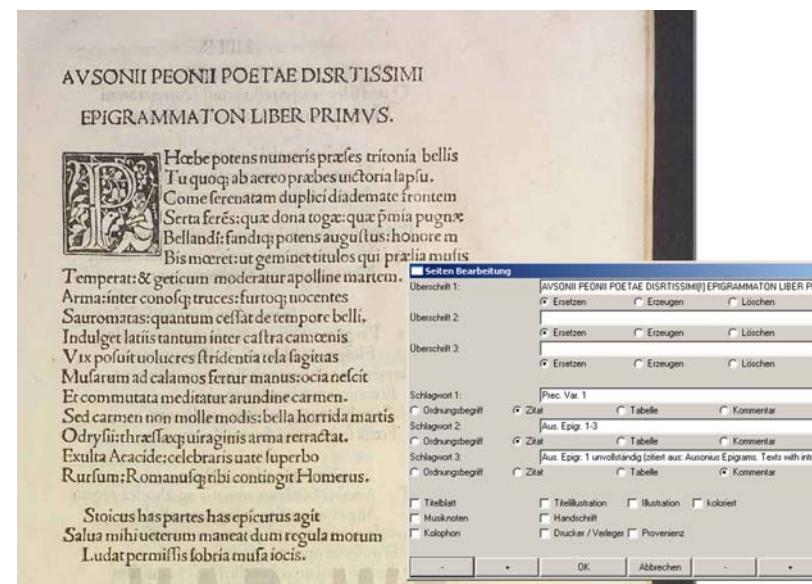


Abbildung 2: Seite der digitalen Ausonius-Ausgabe mit Bearbeitungsfenster des TOC-Editors

```
[...]
<div level="0" >
<head>AVSONII PEONII POETAE DISRTISSIMI[!]
  EPIGRAMMATON LIBER PRIMVS</head>
<div level="1" >
  <page>
```

6 <<http://diglib.hab.de/wdb.php?dir=inkunabeln/12-5-poet-2f-1>> (14.02.2006).

```

    <image extref="00015.jpg" seqno="019"
nativeno="15" />
<text type="reference" >Prec. Var. 1</text>
<text type="reference" >Aus. Epigr. 1-3</text>
<text type="comment" >Aus. Epigr. 1 unvollständig
(zitiert aus: Ausonius Epigrams. Texts with
introduction and commentary by N. M. Kay. -
London 2001)</text>
</page>
  <page>
    <image extref="00016.jpg" seqno="020" na-
tiveno="16" />
<text type="reference" >Aus. Epigr. 5-6, 8-9,
45</text>
<text type="comment" >Eingeschoben: Ignota
aeternae ne sint tibi tempora Romae/ Regibus et
patrum ducta sub imperiis</text>
</page>
  <page>
    <image extref="00017.jpg" seqno="021"
nativeno="17" />
<text type="reference" >Aus. Epigr. 4, 10-
11</text>
<text type="comment" >Eingeschoben: Hactenus
ascripsi fastos: si fors uolet: ultra / Adiiciam:
si non: qui legis: adiicies</text>
<text type="comment" >Eingeschoben: Urbis ab
aeternae deductam rege Quirino / Annorum seriem
cum procule accipies.</text>
</page>
  <page>
    <image extref="00018.jpg" seqno="022"
nativeno="18" />
<text type="reference" >Aus. Epigr. 12-16</text>
</page>
  <page>
    <image extref="00019.jpg" seqno="023"
nativeno="19" />
<text type="reference" >Aus. Epigr. 17-21</text>
</page>
  <page>
    <image extref="00020.jpg" seqno="024"

```

```

nativeno="20" />
<text type="reference" >Aus. Epigr. 22-26</text>
</page>
[...]
```

Das Beispiel verdeutlicht, wie die einzelnen Komponenten zusammenwirken und das Dokument erschließen. Im Folgenden sollen die daraus resultierenden Recherchemöglichkeiten vorgestellt werden.

### Recherchemöglichkeiten innerhalb der vdlb

Aus den im vorangegangenen Kapitel dargestellten Bestandteilen der inhaltlichen Erschließung der Inkunabeln resultiert eine große Daten- und Informationsmenge, die zu jedem einzelnen Digitalisat verfügbar ist. Damit diese in ihrer ganzen Bandbreite genutzt werden können, sollen durch den Einbau von verschiedenen Recherchemöglichkeiten auf der Internetseite des Projektes alle eingetragenen Daten einzeln gesucht werden können.<sup>7</sup> Es soll also möglich sein, sowohl nach den üblichen Formulareintragen als auch nach den Eintragungen der inhaltlichen Erschließung zu suchen. Grundsätzlich sind dafür drei Suchmöglichkeiten auf der Projektseite verfügbar: eine Indexsuche, eine Maskensuche und eine Einfache Suche. Das Suchinstrument der Einfachen Suche folgt „modifizierten Google-regeln“. Die Index- und Maskensuche erlauben die Recherche nach ausgewählten Kategorien, die sich an dem oben formulierten Anspruch orientieren. Ergänzend dazu unterstützen die Suchmaschinen die bereits bestehenden Ressourcen zur Inkunabelkunde.

### Einbindung der vdlb in bestehende Ressourcen

Die vdlb ist eng verbunden mit bereits existierenden Internetressourcen zum Thema Inkunabelkunde. Dementsprechend kam den ISTC-Nummern (Incunabula Short Title Catalogue) eine wichtige Funktion bei der Digitalisierungskampagne zu, denn die ISTC-Titelaufnahmen wurden für alle nachgewiesenen Inkunabeln der Projektpartner in Form einer suchbaren Komponente des Gesamtservers implementiert. So wird nach dem Eintrag

<sup>7</sup> Diese Arbeiten sind zurzeit (Stand: Februar 2006) noch nicht im vollen Umfang abgeschlossen, jedoch wird dies zeitnah geschehen.

der ISTC-Nummern in die XML-Struktur die komplette ISTC-Beschreibung inklusive der umfangreichen bibliografischen Angaben zugewiesen. Sie kann somit jederzeit auf der Projektseite abgerufen werden. Im Gegenzug wurden die digitalisierten Inkunabeln dem ISTC für die weitere Informationsverarbeitung gemeldet. Die vorgestellten Suchmaschinen unterstützen folglich auch die Recherche nach ISTC-Nummern. Des Weiteren kann auch nach den GW-Nummern<sup>8</sup> (Gesamtkatalog der Wiegendrucke) und den BSB-Ink-Signaturen<sup>9</sup> (Bayerische Staatsbibliothek, Inkunabelkatalog) der jeweiligen Inkunabeln gesucht und geordnet werden. Umgekehrt können die Produkte der vdlb über ISTC-online<sup>10</sup> und INKA<sup>11</sup> (Inkunabelkatalog deutscher Bibliotheken) gefunden werden.

Im Laufe des Projektes wurde in Eigenleistung auch eine Neukatalogisierung der Inkunabeln im PICA-OPAC erbracht. Damit verbunden ist der unmittelbare Zugriff auf alle Digitalisate über den Katalog der HAB sowie den GBV möglich. Zusätzlich gibt es einen Linkapparat, in dem bereits bestehende digitalisierte Inkunabeln anderer Internetressourcen und auch digitalisierte Nachschlagewerke angegeben sind.

### Fazit

Wie aus den vorangegangenen Ausführungen ersichtlich wird, versteht sich die vdlb nicht als ein von seiner Umwelt und von weiteren Entwicklungen losgelöstes Projekt, sondern vielmehr wird die Einbindung bzw. Verknüpfung in zukünftige Digitalisierungsvorhaben ausdrücklich erwünscht. Die Anzahl der existierenden Inkunabeln ist überschaubar, außerdem sind sie bibliografisch hervorragend erschlossen, so dass ein bundesweites „Inkunabelportal“ in der mittelbaren Zukunft möglich ist. Die vdlb begreift sich als ein Baustein auf dem Weg dazu. Das Angebot bietet Orientierungsmöglichkeiten, aber sicher auch die Chance der kritischen Auseinandersetzung. Der Grundsatz der vdlb, auf eine inhaltliche Erschließung zu verzichten und stattdessen eine größere Quantität an Digitalisaten zu erreichen, birgt

---

8 <<http://www.gesamtkatalogderwiegendrucke.de/>> (14.02.2006).

9 <<http://mdz1.bib-bvb.de/cocoon/bsbink/start.html>> (14.02.2006).

10 <<http://www.bl.uk/catalogues/istc/index.html>> (14.02.2006).

11 <<http://www.inka.uni-tuebingen.de/>> (14.02.2006).

sicherlich Gefahren in sich, macht aber auf der anderen Seite auch deutlich, dass es möglich ist, innerhalb von verhältnismäßig kurzer Zeit signifikante Bestände zu digitalisieren und dabei die Benutzungsfreundlichkeit nicht außer Acht zu lassen. Es wäre wünschenswert, wenn andere Projekte das Angebot der vdlb aufgreifen und es so in der näheren Zukunft möglich wäre, weltweit auf die bestehenden Inkunabelbestände zuzugreifen. Auch wenn dies zum gegenwärtigen Zeitpunkt noch eher eine Vision denn nahe Realität ist, so haben die Arbeit an dem Projekt und das Ergebnis gezeigt, dass es durchaus die Chance gibt, diese Vision zu verwirklichen.

## Anhang

Zur weiteren Verdeutlichung seien an dieser Stelle zwei Screenshots von digitalisierten Inkunabelseiten mit Illustrationen angeführt, die der digitalen Bibliothek der HAB entstammen.

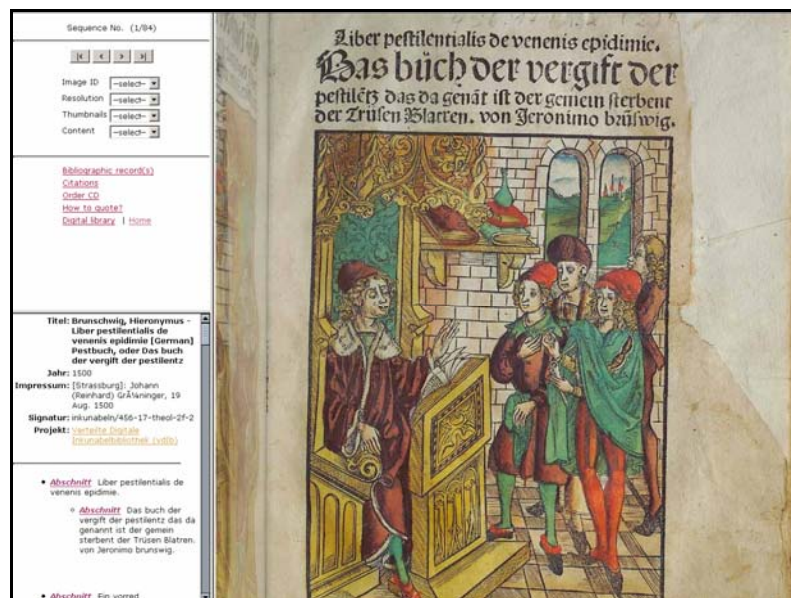


Abbildung I: Titelblatt des Pestbuches von Hieronymus Brunswig<sup>12</sup>

12 Hieronymus Brunswig, Liber pestilentialis de venenis epidimie. Das buch der vergift der pestile[n]tz das da gena[n]t ist der gemein sterbent der Trüsen Blaten. Straßburg 1500. Hier S. 1 der digitalen Ausgabe: <<http://diglib.hab.de/wdb.php?dir=inkunabeln/456-17-theol-2f-2>> (14.02.2006).



Abbildung II: Der Anfang des Kapitels über unnütze Bücher im Narrenschiff des Sebastian Brant<sup>13</sup>

\*\*\*

Timo Steyer ist Wissenschaftlicher Mitarbeiter im Forschungsprojekt der „Verteilten Digitalen Inkunabelbibliothek“ an der Herzog August Bibliothek Wolfenbüttel. Ein weiterer Interessenschwerpunkt des Autors liegt in der Erforschung von Ausdrucksformen der spätmittelalterlichen und frühneuzeitlichen Volksfrömmigkeit. E-Mail: [Timo.Steyer@gmx.de](mailto:Timo.Steyer@gmx.de)

13 Sebastian Brant, Stultifera Navis Narragonice p[ro]fect[i]onis nunq[ua]m satis laudata Navis, Basel 1498. Hier S. 21 der digitalen Ausgabe: <<http://diglib.hab.de/wdb.php?dir=inkunabeln/30-4-poet-2>> (14.02.2006).

ZEITSCHRIFTEN DER AUFKLÄRUNG IM NETZ – RETROSPEKTIVE  
DIGITALISIERUNG WISSENSCHAFTLICHER REZENSIONSORGANE  
UND LITERATURZEITSCHRIFTEN

von Sabine Rahmsdorf

*Mit dem an der Universitätsbibliothek Bielefeld durchgeführten und von der DFG geförderten Projekt „Retrospektive Digitalisierung wissenschaftlicher Rezensionsorgane und Literaturzeitschriften des 18. und 19. Jahrhunderts aus dem deutschen Sprachraum“ ist eine im Internet frei zugängliche Sammlung der vollständigen Textkorpora wichtiger Rezensionsorgane und Literaturzeitschriften der deutschen Aufklärung entstanden. Verfügbar sind bisher 45 Zeitschriften, die auf der Grundlage von Mikroverfilmungen des Georg Olms Verlags digitalisiert wurden. Die inhaltliche Erschließung wird weitgehend durch den an der Akademie der Wissenschaften zu Göttingen erstellten „Index deutschsprachiger Zeitschriften 1750-1815“ geleistet, der in das Projekt eingeflossen ist. Die Verbindung von digitaler Reproduktion der Zeitschriften mit hochwertigen inhaltserschließenden Daten eröffnet der Forschung neue Rezeptionsmöglichkeiten. Der Beitrag stellt die entstandene digitale Sammlung mit ihren Recherchemöglichkeiten vor und berichtet über die im Projektverlauf gewonnenen Erfahrungen mit der Bereitstellung und Erschließung historischer Quellen in digitaler Form.*

\*\*\*

Die Zeitschrift ist ein zentrales Kommunikationsmedium der Epoche der Aufklärung. Dies wird dokumentiert in einem vielfältigen Spektrum literarischer Zeitschriften und Rezensionsorgane, gelehrter oder historisch-politischer Journale, das sich im 18. und frühen 19. Jahrhundert im deut-

schen Sprachraum (und natürlich nicht nur hier) entfaltet.<sup>1</sup> Neben langjährigen und überregional bedeutsamen Zeitschriften stehen viele kurzlebige Unternehmungen oder solche mit nur einem lokal eng begrenzten Rezipientenkreis. Die für die Epoche der Aufklärung so kennzeichnende Blüte der Zeitschriften ist dabei Katalysator und Ergebnis einer sich konstituierenden gelehrten bürgerlichen Öffentlichkeit, die für ihr Bedürfnis nach öffentlicher Diskussion und Information in der Zeitschrift mit ihrer inhaltlich flexiblen Form und ihrer durch die periodische Erscheinungsweise angestrebten Aktualität das ideale Medium fand. Gerade die so dezentrale „Gelehrtenrepublik“ in den deutschen Ländern konnte in den literarischen und gelehrten Zeitschriften die wesentlichen Diskurse der Epoche führen.<sup>2</sup> Die Bedeutung der Zeitschriften für die Erforschung des Jahrhunderts der Aufklärung sowie der angrenzenden Epoche liegt daher seit langem auf der Hand. Neben die Analyse einzelner Zeitschriftenbeiträge tritt dabei die Betrachtung des Mediums selbst, denn die Untersuchung der Beziehungen zwischen Herausgebern/innen, Beiträgern/innen und Rezipienten/innen verschafft Aufschluss über persönliche Verbindungen und bestehende Netzwerke zwischen den Trägern der Aufklärungsdiskussion und der daran Anteil nehmenden Öffentlichkeit. Vergleichende Untersuchungen der oft genug aufeinander Bezug nehmenden Beiträge in den verschiedensten Zeitschriften können wertvolle Erkenntnisse über den Verlauf einzelner Debatten und Diskursthemen und ihrer Rezeption erbringen.<sup>3</sup>

- 
- 1 Einen fundierten Überblick bietet Kirchner, Joachim, Die Zeitschriften des deutschen Sprachgebietes von den Anfängen bis 1830 (= Ders. (Hg.), Bibliographie der Zeitschriften des deutschen Sprachgebietes bis 1900, Bd. 1, Stuttgart 1969).
  - 2 Einführend zur Thematik siehe Wilke, Jürgen, Literarische Zeitschriften des 18. Jahrhunderts (1688-1789), Stuttgart 1978; Hocks, Paul; Schmidt, Peter, Literarische und politische Zeitschriften 1789-1805. Von der politischen Revolution zur Literaturrevolution, Stuttgart 1975.
  - 3 Auf den Stand der Forschung kann an dieser Stelle nicht eingegangen werden, beispielhaft seien einige neuere Arbeiten zur Thematik angeführt: Schneider, Ute, Friedrich Nicolais Allgemeine Deutsche Bibliothek als Integrationsmedium der Gelehrtenrepublik, (Mainzer Studien zur Buchwissenschaft 1), Wiesbaden 1995; Heinz, Andrea (Hg.), „Der Teutsche Merkur“ – die erste deutsche Kulturzeitschrift?, (Ereignis Weimar-Jena. Kultur um 1800 2), Heidelberg 2003; Matuschek, Stefan (Hg.), Organisation der Kritik. Die Allgemeine Literatur-

Der Bedeutung der Zeitschriften für die Aufklärungsforschung steht die Zugänglichkeit dieses Quellenmaterials entgegen. Das in seinem Erhaltungszustand fragile Material findet sich als historisch gewachsener und verstreuter Besitz in Bibliotheken und ist für die Forschung im Original nur erschwert zugänglich. Insbesondere bei vergleichenden Fragestellungen stellt die Arbeit mit Originalen an verschiedenen Standorten eine Erschwernis dar. Daher hat es in der Vergangenheit Bemühungen gegeben, dieses Quellenmaterial besser zu erschließen. Das Unternehmen „Index deutschsprachiger Zeitschriften 1750-1815“<sup>4</sup> der Akademie der Wissenschaften zu Göttingen, von 1976 bis 1987 gefördert durch die Stiftung Volkswagenwerk und die Deutsche Forschungsgemeinschaft (DFG), hatte die „Erschließung der in den Zeitschriften der deutschen Aufklärung sachlich diskutierten Themen und dichterisch behandelten Motive in repräsentativer und dennoch möglichst breiter Auswahl“<sup>5</sup> zum Ziel. Nicht erfasst wurden die moralischen Wochenschriften und Rezensionsorgane. Im Ergebnis konnten 195 Zeitschriften mit fast 100.000 Beiträgen erschlossen werden, dies entspricht „etwa 20% der in dieser Zeit publizierten Inhalte“<sup>6</sup>. Alle in den ausgewerteten Zeitschriften enthaltenen Beiträge einschließlich Rezensionen wurden bibliografisch erfasst, Autoren/innen und beteiligte Personen anonymer Beiträge wo möglich ermittelt und inhaltserschließende Schlagwörter vergeben. Ein weiteres Großprojekt ist die von der Kulturstiftung der Länder herausgegebene und vom Georg Olms Verlag verlegte Sammlung „Deutsche Zeitschriften des 18. und 19. Jahrhunderts“, die über 500 Zeitschriften als Mikrofiche-Volltextverfilmung umfasst.<sup>7</sup>

In den vergangenen rund zehn Jahren ist mit der Etablierung des In-

---

Zeitung in Jena 1785-1803, (Ereignis Weimar-Jena. Kultur um 1800 5), Heidelberg 2004.

- 4 Akademie der Wissenschaften zu Göttingen (Hg.), Index deutschsprachiger Zeitschriften 1750-1815. Erstellt durch eine Arbeitsgruppe unter Leitung von Klaus Schmidt, Hildesheim 1990.
- 5 Ebd., Begleitheft S.VII.
- 6 Ebd., Begleitheft S.VIII.
- 7 Kulturstiftung der Länder (Hg.), Deutsche Zeitschriften des 18. und 19. Jahrhunderts, Hildesheim 1994ff.

ternet als Informations- und Kommunikationsmittel auch der Wissenschaft die digitale Reproduktion historischer Literatur und anderen Quellenmaterials und dessen Veröffentlichung als Onlinepublikation zu einer wichtigen Möglichkeit geworden, auch schwer zugängliches Quellenmaterial prinzipiell weltweit zugänglich zu machen. Bibliotheken – auf kommerzielle Produkte auf diesem Sektor soll an dieser Stelle nicht eingegangen werden – stellen in verschiedenen Digitalisierungsprojekten wichtige Teile ihrer historischen Bestände als digitale Reproduktionen einer interessierten Öffentlichkeit frei im Internet zur Verfügung.<sup>8</sup> In Deutschland hat die DFG durch ihr Förderprogramm „Retrospektive Digitalisierung von Bibliotheksbeständen“, das 2004 in dem neuen Förderprogramm „Kulturelle Überlieferung“ aufgegangen ist, maßgeblich zur Initiierung und Finanzierung dieser Aktivitäten beigetragen. Mit dem Aufbau von zwei nationalen Digitalisierungszentren, dem Göttinger Digitalisierungszentrum (GDZ) der Niedersächsischen Staats- und Universitätsbibliothek Göttingen und dem Münchner Digitalisierungszentrum an der Bayerischen Staatsbibliothek München, wurden zwei Kompetenz- und Servicezentren für retrospektive Digitalisierung von Bibliotheksbeständen geschaffen. Auf den Webseiten der beiden Digitalisierungszentren finden sich jeweils Übersichten über abgeschlossene und laufende Digitalisierungsprojekte.<sup>9</sup> Um die Vielzahl der verschiedenen Digitalisierungsprojekte zusammenzuführen und einen zentralen Zugriff auf die in digitalisierter Form vorliegenden Drucke und anderes Quellenmaterial zu ermöglichen, befindet sich zurzeit, gefördert wiederum von der DFG, ein Zentrales Verzeichnis digitalisierter Drucke im Aufbau. Beteiligt sind hieran die AG Sammlung Deutscher Drucke (SDD), die Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG) und das Hochschulbibliothekszentrum (HBZ) des Landes Nord-

---

8 Eines der international bekanntesten und bedeutendsten Digitalisierungsprojekte ist die von der französischen Nationalbibliothek erstellte Sammlung gallica, <<http://gallica.bnf.fr/>> (16.02.2006).

9 Göttinger Digitalisierungszentrum (GDZ), Niedersächsische Staats- und Universitätsbibliothek Göttingen, <<http://gdz.sub.uni-goettingen.de/de/index.html>> (21.02.2006); Münchner Digitalisierungszentrum, Bayerische Staatsbibliothek München, <<http://www.bsb-muenchen.de/mdz/>> (21.02.2006).

rhein-Westfalen. Eine erste Testversion ist im Internet verfügbar.<sup>10</sup>

Die Vorteile einer Bereitstellung digitalisierter Quellen oder historischer Buchbestände im World Wide Web liegen auf der Hand. Neben der verbesserten Zugänglichkeit kann durch die Erstellung virtueller Kompendien von Quellen nicht nur im Original verstreutes, aber thematisch zusammengehöriges Material zusammengeführt werden, durch übergreifende Suchmöglichkeiten lassen sich Einzelwerke und Sammlungen zudem in neuartiger Weise inhaltlich erschließen und können so auch die Bearbeitung neuer Fragestellungen ermöglichen. Einschränkend muss jedoch gesagt werden, dass digitalisierte historische Dokumente überwiegend noch als digitales Image, also als Bilddatei der einzelnen Seiten eines Dokuments angeboten werden. Die inhaltliche Erschließung digitaler Quellen konzentriert sich damit aber noch ganz überwiegend auf die Möglichkeit der Stichwortsuche in zusätzlich maschinenlesbar erfassten Inhaltsverzeichnissen und Registern oder in der Suche nach ergänzend angebotenen Schlagwörtern und Thesauri. Die Konvertierung der als Image vorliegenden Seiten in einen maschinenlesbaren Volltext mit Hilfe von Optical Character Recognition-Software (OCR) stößt bei historischen Quellen und Dokumenten aufgrund der verwendeten nicht normierten Schriften, Schreibweisen oder Abkürzungen nach wie vor auf große Schwierigkeiten. Genau dies ist aber die Voraussetzung, um eine unkomplizierte Möglichkeit der Volltextsuche zu ermöglichen, da sich zumal bei öffentlich geförderten Projekten die grundsätzlich denkbare aber kostenintensive Texterfassung des gesamten Inhalts eines digitalisierten Dokuments zumeist verbietet. In diesen Kontext bibliotheksseitiger Digitalisierungsprojekte ist auch das hier vorzustellende Projekt der Universitätsbibliothek Bielefeld einzuordnen.

### Das Projekt

Das an der Universitätsbibliothek Bielefeld von 2000 bis 2003 durchgeführte Projekt „Retrospektive Digitalisierung wissenschaftlicher Rezensi-

10 Zentrales Verzeichnis digitalisierter Drucke, <<http://www.digitalisiertedrucke.de/>> (16.02.2006).

onsorgane und Literaturzeitschriften des 18. und 19. Jahrhunderts aus dem deutschen Sprachraum“ (Phase 1)<sup>11</sup> hatte die digitale Rekonstruktion der vollständigen Korpora zentraler wissenschaftlicher Rezensionsorgane und Literaturzeitschriften der deutschen Aufklärung und folgender Epochen zum Ziel. Berücksichtigt wurden Zeitschriften aus der zweiten Hälfte des 18. und aus dem frühen 19. Jahrhundert. Die digitalisierten Quellen sollten frei zugänglich über das Internet zur Verfügung gestellt werden und durch Suchmöglichkeiten sowie durch ein Browsing erschlossen werden. Das Projekt wurde von der DFG im Rahmen des Schwerpunkts „Retrospektive Digitalisierung von Bibliotheksbeständen“ gefördert. Im Ergebnis stehen 45 Zeitschriften zur Verfügung mit der Möglichkeit zur Recherche in insgesamt 82.000 Artikeln bzw. 460.000 Seiten. Digitalisiert wurden zum Beispiel die von Friedrich Nicolai herausgegebene Allgemeine deutsche Bibliothek, die Bibliothek der schönen Wissenschaften und der freyen Künste, das Deutsche Museum / Neues Deutsches Museum oder der Teutsche Merkur / Neue Teutsche Merkur.<sup>12</sup>

Das Projekt wurde durchgeführt in Kooperation mit dem Georg Olms Verlag und mit der Akademie der Wissenschaften zu Göttingen. Die Digitalisierung der Zeitschriften erfolgte nicht anhand der Originale, sondern auf der Grundlage von Mikrofiches aus der von der Kulturstiftung der Länder herausgegebenen Mikrofiche-Edition „Deutsche Zeitschriften des 18. und 19. Jahrhunderts“, die vom Olms Verlag zur Verfügung gestellt wurden. Die recherchierbaren Daten wiederum sind Ergebnis der Erschließungsarbeiten des Unternehmens „Index deutschsprachiger Zeitschriften 1750-1815“ der Akademie der Wissenschaften zu Göttingen. Damit ist es gelungen, die wertvollen Erschließungsdaten des „Index“ nachzunutzen und in neuer Form zu präsentieren und durch die Verbindung mit den Volltexten der Zeitschriften einen Mehrwert in dem digitalen Rechercheangebot zu erzielen.

11 Retrospektive Digitalisierung wissenschaftlicher Rezensionsorgane und Literaturzeitschriften des 18. und 19. Jahrhunderts aus dem deutschen Sprachraum, <<http://www.ub.uni-bielefeld.de/diglib/aufklaerung/>> (16.02.2006).

12 Eine Auflistung der zur Verfügung stehenden Zeitschriften ist unter der oben angegebenen URL des Projekts einsehbar.

## Das Rechercheangebot

Zur Navigation und Recherche in den digitalisierten Zeitschriften werden sowohl Möglichkeiten zum Browsing als auch eine Suchoberfläche angeboten. Beim Browsing kann ausgehend von der Auflistung aller Zeitschriften eine Übersicht über die Bände einer Zeitschrift und auf der nächsten Ebene über die Inhalte des Einzelbandes angesteuert werden. Ausgehend vom Eintrag eines Beitrags im Inhaltsverzeichnis kann der zugehörige Volltext aufgerufen werden. Innerhalb der digitalisierten Dokumente kann Seite für Seite und in Schritten von fünf bzw. zehn Seiten vorwärts und rückwärts geblättert werden, auf diese Weise kann ein kompletter Zeitschriftenband auch Seite für Seite betrachtet werden.

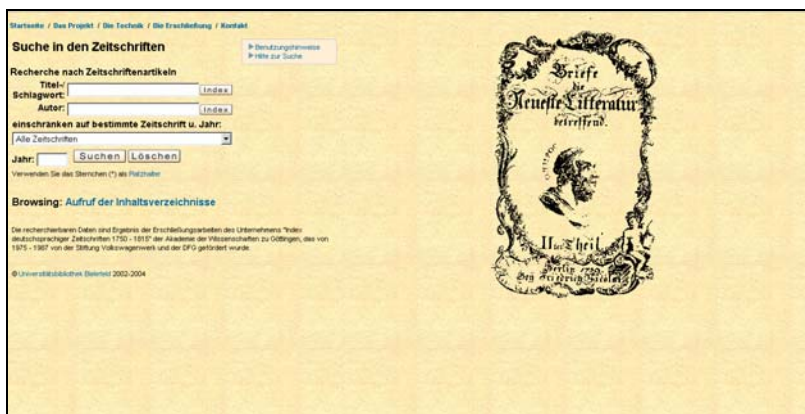


Abbildung 1: Die Suchoberfläche mit Möglichkeit zum Browsing über die Zeitschriften

Die Suchoberfläche bietet die Möglichkeit der personenbezogenen oder thematischen Suche nach einzelnen Zeitschriftenbeiträgen. Möglich ist die Suche nach dem Autor bzw. der Autorin eines Beitrags oder einer Rezension sowie nach sonstigen an der Erstellung eines Beitrags beteiligten Personen, wie beispielsweise einem Übersetzer oder einer Übersetzerin. Das Suchfeld „Titel-/Schlagwort“ ermöglicht die Suche nach Stichwörtern

aus dem Titel eines Beitrags und nach inhaltserschließenden Schlagwörtern, wie sie im „Index“ vergeben wurden. Die Schlagwörter bezeichnen zum Beispiel geografische Bezüge oder sind Personen- oder Sachschlagwörter. Jede Suchanfrage kann entweder über den gesamten Zeitschriftenbestand erfolgen oder auf eine bestimmte Zeitschrift und / oder bestimmte Erscheinungsjahre eingegrenzt werden. Als Ergebnis einer Recherche wird eine Kurztitelliste der Treffer angezeigt, aus der heraus die Vollanzeige jedes Treffers mit den vollständigen bibliografischen Angaben des Beitrags einschließlich Schlagwörter und dem Link zum zugehörigen Volltext aufgerufen werden kann. Die Recherche in den digitalisierten Zeitschriften erfolgt auf der Basis der erschließenden Metadaten, also anhand der bibliografischen Angaben einschließlich inhaltserschließender Schlagwörter zu jedem einzelnen Zeitschriftenbeitrag. Eine Suche in den Volltexten der Zeitschriften wird bisher nicht angeboten.

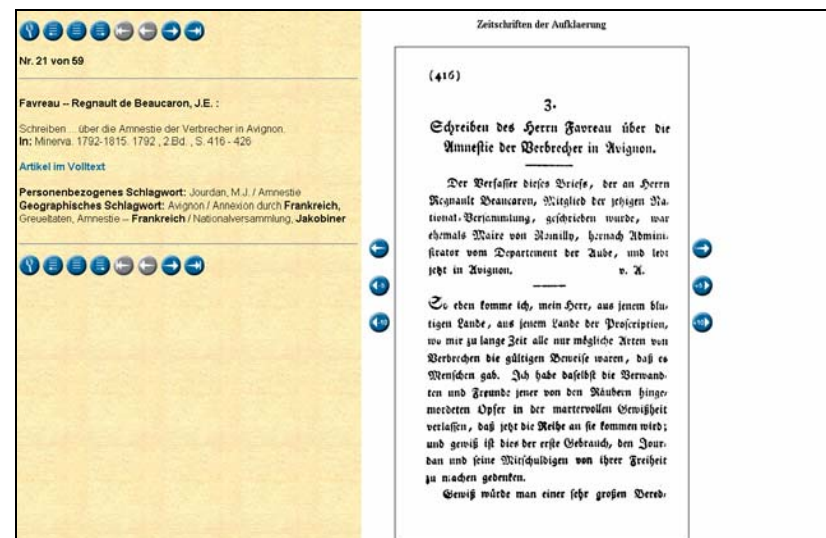


Abbildung 2: Trefferanzeige mit zugehörigem Volltext und Funktion zum Blättern im Dokument

### Die technische Realisierung

Zur technischen Realisierung des Projekts waren verschiedene Schritte erforderlich, die im Folgenden skizziert werden sollen.

Scannen vom Mikrofiche und Speicherung von Einzelseiten als Bitmap-Dateien:

- Die technischen Parameter für das Scannen der vom Olms Verlag bereitgestellten Masterfiches durch einen externen Dienstleister folgten den üblichen und von der DFG empfohlenen Vorgaben<sup>13</sup>: Die Scans der Einzelseiten wurden bitonal (schwarz/weiß), mit 600 dpi Auflösung als Bitmap-Dateien im TIFF-Level-4-Format erstellt. Das Tagged Image File Format (TIFF) ist für die langfristige Speicherung von Bilddaten in der Regel das zu empfehlende Format, das zudem eine hohe Bildqualität gewährleistet und als Ausgangsformat für die Erstellung komprimierter Bildformate genutzt werden kann, die für die schnelle Datenübertragung über das Internet eher geeignet sind. Die Benennung der Bitmap-Dateien erfolgte rein numerisch, entsprechend der Reihenfolge der Originalseiten auf den vom Olms Verlag bereitgestellten Mikrofiches.
- **Aufbereitung der inhaltserschließenden Daten:**  
Um die von der Akademie der Wissenschaften zu Göttingen als Textdateien bereitgestellten inhaltserschließenden Daten aus dem „Index deutschsprachiger Zeitschriften 1750-1815“ in einen durchsuchbaren Datenpool zu überführen und für die Einspielung in eine Datenbank vorzubereiten, wurden diese in das als internationaler Standard vorliegende Auszeichnungsformat SGML (Standard Generalized Markup Language) konvertiert. Für fünf Zeitschriften lagen keine oder nur unvollständige Erschließungsdaten aus dem „Index“ vor, darunter die Allgemeine deutsche Bibliothek. Hier mussten ersatzweise Inhaltsverzeichnisse bzw. Register auf der Grundlage geeigneter Erfassungsre-

13 Die aktuelle Version der „Praxisregeln im Förderprogramm ‚Kulturelle Überlieferung‘“ der DFG, die in erster Linie technische Hinweise zu den Bereichen Verfilmung und Digitalisierung geben, ist verfügbar unter: [http://www.dfg.de/forschungsfoerderung/formulare/download/12\\_151.pdf](http://www.dfg.de/forschungsfoerderung/formulare/download/12_151.pdf) (16.02.2006).

geln im Double-Key-Verfahren<sup>14</sup> manuell erfasst und ebenfalls nach SGML konvertiert werden. Anschließend wurden alle in SGML vorliegenden Erschließungsdaten mit Hilfe von Perl-Skripten zur Einspielung in das eingesetzte Datenbanksystem BRS/Search (Bibliographic Retrieval System) konvertiert, mit dem zahlreiche Datenbanken der Universitätsbibliothek Bielefeld aufbereitet werden. Zu jedem Datensatz wurde die erste Seite des zugehörigen Aufsatzes entsprechend der Originalseitenzählung gespeichert. Die manuelle Erfassung von Inhaltsverzeichnissen und Registern und die Konvertierung nach SGML sowohl der Daten des „Index deutschsprachiger Zeitschriften 1750-1815“ wie auch der Daten der manuellen Erfassung war Bestandteil der Arbeiten eines externen Dienstleisters.

- **Erzeugung der Paginierungskonkordanz:**  
Die Paginierungskonkordanz dient der Sicherstellung des seitenrichtigen Zugriffs von den inhaltserschließenden SGML-Dateien (Metadaten) via Hyperlink auf die jeweils zugehörigen TIFF-Dateien (Dokumentenvolltexte). Sie beinhaltet die exakte Zuordnung zwischen der rein numerischen Benennung der TIFF-Dateien einerseits und der Seitenzählung der Originale andererseits. Die Paginierungskonkordanz wurde für jeden Einzelband mittels intellektueller Prüfung erstellt.
- **Generierung der Benutzeroberfläche:**  
Bei der Extraktion der in SGML vorliegenden Artikel-Metadaten mittels Perl-Skript für den Aufbau der BRS/Search-Datenbank wurden neben dem Datenbankeingabeformat statische, miteinander verknüpfte HTML-Seiten erzeugt, welche die Titel-, Band- und Heftstruktur widerspiegeln.  
Die Web-basierte Benutzeroberfläche besteht aus statischen HTML-Eingangsseiten, die inhaltliche Informationen bereithalten und mit den

14 Beim Double-Key-Verfahren wird eine Vorlage von zwei Datentypisten unabhängig voneinander elektronisch erfasst (abgeschrieben) und der erfasste Text anschließend elektronisch gegeneinander abgeglichen. Durch diesen Abgleich können die bei jeder Datenerfassung auftretenden individuellen Fehler durch falsches Abschreiben minimiert und eine hohe Erfassungsgenauigkeit erzielt werden. Double-Keying ist ein anerkanntes Verfahren.

Funktionen für Datenbanksuche und Browsing in der bibliografischen Zeitschriftenstruktur verknüpft sind. Auf diesen Seiten und in der Datenbankergebnisanzeige werden Links auf ein CGI-basiertes Perlskript<sup>15</sup> verankert, das als Parameter Pfad und Benennung der Startdatei erhält. Dieses Skript übernimmt die On-The-Fly-Umwandlung der als Speicherformat, nicht aber für die Darstellung der digitalisierten Dokumente im Internet geeigneten TIFF-Dateien in das WWW-gerechte Grafik-Austausch-Format GIF (Graphics Interchange Format) und steuert gleichzeitig die Navigation über die einzelnen Seiten der Dokumente. In der Benutzeroberfläche wird in einer Framestruktur im linken Fenster die Suche und Anzeige in der BRS/Search-Datenbank oder die Browsing-Navigation in der Struktur der Zeitschriften angeboten, im rechten Fenster erscheint die jeweils angewählte zugehörige Imagedatei mit der Darstellung des ausgewählten Zeitschriftenbeitrags.

Für jede Zeitschrift wurde eine eigene Frontdoor-Seite erstellt, die neben den bibliografischen Daten der Zeitschrift den Einstieg zu den HTML-Seiten mit der Übersicht über die Bandstruktur und den Inhaltsverzeichnissen der Einzelbände sowie zu den zugehörigen Volltexten der enthaltenen Artikel bietet. Die URL der Frontdoor-Seite wird auch in der Zeitschriftendatenbank (ZDB) und der Elektronischen Zeitschriftenbibliothek (EZB) nachgewiesen, so dass die digitalisierten Zeitschriftentitel auch in diesen überregionalen Katalogen recherchierbar sind. Über Internetsuchmaschinen sind die digitalisierten Zeitschriftentitel ebenfalls suchbar.

### **Blick zurück: Erfahrungen aus dem Projektverlauf**

Nach Abschluss des Projekts ist es angebracht, aus den gewonnenen Erfahrungen mit der digitalen Bereitstellung eines historischen Quellenkorpus, wie es die zur Verfügung gestellten Zeitschriften der Aufklärung darstellen, ein Resümee zu ziehen. Insgesamt haben sich die gewählten technischen Vorgehensweisen und Arbeitsabläufe mit Blick auf Praktikabi-

<sup>15</sup> Der Standard CGI (Common Gateway Interface) dient, allgemein gesprochen, der Ausführung externer Programme auf einem WWW-Server.

lität und erzieltes Ergebnis bewährt, einige Punkte jedoch illustrieren, welche besonderen Anforderungen und Schwierigkeiten die digitale Reproduktion gerade historischer Dokumente immer noch in sich birgt:

Das gewählte Verfahren der Digitalisierung vom Mikrofiche und nicht vom Original stellte im Kontext dieses Projekts ein sinnvolles Vorgehen dar, da auf diese Weise davon abgesehen werden konnte, die an verschiedenen Bibliotheksstandorten verfügbaren Originale für eine Digitalisierung mit hohem Aufwand erneut zusammenzustellen. Das Scannen von einer Sekundärform der Originalquelle erfordert jedoch ein besonders hohes Maß an Qualitätskontrolle für alle produzierten Imagedateien, um eine einwandfreie Präsentation zu gewährleisten. Entsprechend hoch war der Personal- und Zeitaufwand hierfür. Die Digitalisierung vom Mikrofiche ist daher in anderen Kontexten sorgfältig gegen andere Faktoren abzuwägen und kann nicht etwa generell als Alternative zum Scannen des Originals bezeichnet werden.

Der Workflow von der Digitalisierung der Mikrofiches bis zur Präsentation im Internet beinhaltete einen relativ hohen Anteil an manueller und intellektueller Bearbeitung: Für die manuelle Erfassung von Inhaltsverzeichnissen bzw. Registern der nicht im „Index“ erfassten Zeitschriften mussten genaue Erfassungsregeln definiert werden, die bereits die Logik der anschließenden SGML-Konvertierung berücksichtigen mussten. Dies war in besonderem Maße erforderlich, weil die manuelle Erfassung quasi „mechanisch“ durch Schreibkräfte ohne spezifische Sach- oder Sprachkenntnisse vorgenommen wurde und nicht etwa durch speziell geschultes Personal. Aufgrund der hohen Variabilität bei der Gestaltung der Inhaltsverzeichnisse und Register, beispielsweise hinsichtlich der verwendeten Abkürzungen und Verweisungszeichen, mussten diese Erfassungsregeln mit entsprechendem Personalaufwand jedoch jeweils für jede Zeitschrift und teilweise auch für jeden einzelnen Band individuell erstellt werden.

Die Erstellung der Paginierungskonkordanz als notwendigem Bindeglied zwischen inhaltserschließenden Metadaten und zugehörigem Volltext erfolgte durch den intellektuellen Abgleich zwischen den in den Metadaten enthaltenen Seitenzahlen, der Originalseitenzählung innerhalb der TIFF-Dateien und der numerischen Benennung der TIFF-Dateien, für

den studentische Hilfskräfte eingesetzt wurden. Trotz des verhältnismäßig hohen Personalaufwands für manuelle Erfassung, Erstellung von Erfassungsregeln und Erstellung der Paginierungskonkordanz zeigte sich im Projektverlauf, dass eine automatisierte Bearbeitung keine gangbare Alternative darstellte, wenn nicht deutliche Qualitätsabstriche in Kauf genommen werden sollten. Die denkbare Erfassung von Inhaltsverzeichnissen und Registern der nicht im „Index“ enthaltenen Zeitschriften durch Digitalisierung der entsprechenden Seiten und anschließende Konvertierung in maschinenlesbaren Volltext mithilfe von OCR war auf Grund zu hoher Fehlerraten bei der Schrifterkennung nicht möglich.

OCR-Software kam im Projekt bisher nicht zum Einsatz, da die Fehlerrate bei der Schrifterkennung als nicht tolerabel erschien und einen zu hohen Aufwand für Fehlerkontrolle und –korrektur bedeutet hätte. Die in den Zeitschriften der Aufklärung verwendeten Frakturschriften waren bisher generell kaum erfolgreich mit OCR zu bearbeiten. Erschwerend war zudem die stark variierende Druckqualität der Originale (durchscheinender Satzspiegel, unscharfes Schriftbild). Bei eigenen Tests war noch gegen Ende des Projekts die Erkennungsrate der verwendeten OCR nicht akzeptabel. Es bleibt abzuwarten, ob die neuesten Versionen geeigneter OCR-Software auch Frakturschriften künftig mit überzeugenden Erkennungsraten bearbeiten können. Die Übertragung der gescannten Images in maschinenlesbaren Volltext mit Hilfe von OCR bleibt ein interessantes Ziel, da sie Voraussetzung für die Realisierung einer Volltextrecherche ist. Dies würde ohne Frage eine Verbesserung der Recherchemöglichkeiten bedeuten, etwa bei der Suche nach in den Zeitschriftenbeiträgen behandelten Themen, erwähnten Personen oder Orten, die sich aus dem Titel nicht erschließen und den Rahmen der Erfassung durch notwendigerweise knappe Schlagwörter übersteigen, oder bei der gezielten Suche nach Begriffen für begriffsgeschichtliche Fragestellungen.

Die entstandene digitale Sammlung von 45 Zeitschriften der Aufklärung hat ihr wesentliches Ziel erreicht, denn das Angebot ermöglicht die komfortable und gezielte Recherche nach einzelnen Zeitschriftenbeiträgen. Insbesondere für eine vergleichende Arbeit mit diesen Quellen stellt das zur Verfügung stehende Korpus eine wichtige Erleichterung dar. Die

Sammlung erfährt eine ausgesprochen positive Benutzerresonanz, wie sich immer wieder an Kommentaren, die uns per E-Mail erreichen, zeigt. Für die positive Aufnahme spricht auch die rege Nutzung des Angebots, das in der Nutzungsstatistik des gesamten Webangebots der Universitätsbibliothek Bielefeld stets vordere Plätze belegt und circa 60.000 bis 80.000 Zugriffe (=Seitenbewegungen) pro Monat verzeichnet. Dabei überwiegt der Anteil externer Zugriffe, dies ist insbesondere auf die Recherchierbarkeit der Zeitschriften in Internetsuchmaschinen zurückzuführen.

### **Blick nach vorn: Fortsetzung des Projekts**

Die Fortsetzung des Projekts wurde inzwischen durch die DFG bewilligt. Zielsetzung dieser zweiten Projektphase, in der die Kooperation mit dem Georg Olms Verlag und der Akademie der Wissenschaften zu Göttingen fortgesetzt wird, ist die digitale Bereitstellung aller übrigen im „Index deutschsprachiger Zeitschriften 1750-1815“ erschlossenen 150 Zeitschriften. Dabei werden die bisher eingesetzten technischen Verfahren und Workflows im Wesentlichen beibehalten. Ein weiterer Kooperationspartner wird die Klassikstiftung Weimar, Herzogin Anna Amalia Bibliothek sein, die ein eigenes DFG-Projekt zur Digitalisierung und Erschließung von Zeitschriften des Weimar-Jenaer-Literaturkreises um 1800 durchführen und die erstellten digitalen Reproduktionen in das Bielefelder Angebot der „Zeitschriften der Aufklärung“ einbringen wird.

\*\*\*

*Dr. Sabine Rahmsdorf hat mit einer Arbeit über Stadt und Architektur in der literarischen Utopie der Frühen Neuzeit im Fach deutsche Literaturwissenschaft an der Universität Hannover promoviert. Sie ist seit 2001 an der Universitätsbibliothek Bielefeld beschäftigt, wo sie nach der Mitarbeit in verschiedenen Projekten die Fachreferate Geschichte und Kunstgeschichte betreut. E-Mail: sabine.rahmsdorf@uni-bielefeld.de*

DIE DIGITALISIERTEN „JAHRESBERICHTE FÜR DEUTSCHE  
GESCHICHTE“ AUS DER ZWISCHENKRIEGSZEIT – EIN SPIEGEL DER  
GESCHICHTSWISSENSCHAFT ZWISCHEN 1925 UND 1938

von Daniel Schlögl

*Die Jahresberichte für deutsche Geschichte der Zwischenkriegszeit (Berichtsjahre 1925-1938) sind für die heutige Forschung in mehrfacher Hinsicht bedeutsam: Mit ihren bibliografischen Nachweisen und Forschungsberichten dokumentieren sie einerseits die zeitgenössische Fachliteratur. Zudem wird aus den Vorworten, dem Wandel der inhaltlichen Gliederung und insbesondere aus den Forschungsberichten die zunehmende Beeinflussung durch politische und ideologische Motivationen deutlich erkennbar. In einzigartiger Weise spiegeln die älteren Jahresberichte somit die Entwicklung der Historiografie in der Zeit der Weimarer Republik und des Nationalsozialismus wider. Die Bände wurden in einem DFG-Projekt digitalisiert und liegen, nach einer Überarbeitung, in einer XML-Datenbank vor.*

\*\*\*

Die *Jahresberichte für deutsche Geschichte* können auf eine über 125jährige Entwicklung zurückblicken und sind damit eine der traditionsreichsten Einrichtungen der geisteswissenschaftlichen Informationsinfrastruktur in Deutschland.<sup>1</sup> Neben die konventionellen gedruckten Jahresbibliografien<sup>2</sup> ist inzwischen die akkumulierende Online-Datenbank als Hauptpublikati-

onsform getreten; sie stellt mit derzeit etwa 290.000 Titeleinträgen ab dem Berichtsjahr 1985 die umfassendste Fachbibliografie zu allen Epochen und Sachgebieten der deutschen Geschichte dar.<sup>3</sup>

Die Umstellung auf eine EDV-gestützte Arbeitsweise erfolgte zu Beginn der 1990er Jahre zeitgleich mit der Integration der *Jahresberichte* als Langzeitvorhaben in die neu konstituierte Berlin-Brandenburgische Akademie der Wissenschaften. Von Beginn an wurde es als ein wichtiges Ziel angesehen, nicht nur die aktuelle Literaturproduktion zu dokumentieren, sondern auch die älteren Druckbände retrospektiv in die digitale Form zu überführen, um somit sukzessive einen möglichst weit zurückreichenden, vollständigen Literaturnachweis zur deutschen Geschichte anbieten zu können. Für die nach dem Zweiten Weltkrieg erarbeiteten Jahressbände der Bibliografie wird dies durch die manuelle Übertragung der Titeleinträge in die Datenbank praktiziert. Die zwischen 1927 und 1940 erschienenen Bände der *Jahresberichte für deutsche Geschichte* wurden hingegen in einem eigenständigen DFG-Projekt digitalisiert. Hierfür war entscheidend, dass die älteren *Jahresberichte* mit ihrer Parallelität von Bibliografien und Forschungsberichten eine grundlegend andere Struktur als die Bände der nach dem Zweiten Weltkrieg begonnenen „Neuen Folge“ aufweisen. Hinzu kamen weitere Beweggründe, für deren Herleitung ein kurzer Blick auf die Geschichte geworfen werden soll.

Die *Jahresberichte* starteten 1880 als *Jahresberichte der Geschichtswissenschaft*.<sup>4</sup> Es handelte sich um ein Referatsorgan mit universalgeschichtlicher Ausrichtung, wobei die jeweils besprochene Literatur in den Fußnoten nachgewiesen wurde. Im Verlauf des Ersten Weltkriegs mussten die Arbeiten an dem Informationsmittel, das inzwischen auch internationale Reputation gewonnen hatte, eingestellt werden. Nach Kriegsende startete ein erster Versuch der Wiederaufnahme, wobei nicht zuletzt aufgrund der finanziellen und personellen Engpässe und der abgerissenen internationalen Kontakte eine inhaltliche Einengung erfolgte, die sich im neuen Titel

1 Vgl. Winfried Enderle, *Geschichtswissenschaft, Fachinformation und das Internet*, in: eForum zeitGeschichte 3/4, 2001 <[http://www.eforum-zeitgeschichte.at/3\\_01a7.pdf](http://www.eforum-zeitgeschichte.at/3_01a7.pdf)> (25.10.2006), S. 3.

2 Zuletzt erschienen: Berlin-Brandenburgische Akademie der Wissenschaften (Hg.), *Jahresberichte für deutsche Geschichte, Neue Folge*, 56 (2004), Berlin 2005.

3 Homepage: <<http://www.jdg-online.de/>> (25.10.2006).

4 Vgl. Historische Gesellschaft zu Berlin (Hg.) *Jahresberichte der Geschichtswissenschaft*, 36 Bände, Berichtsjahre 1878-1913, Berlin 1880-1916.

*Jahresberichte der deutschen Geschichte* widerspiegelt.<sup>5</sup> Das Unternehmen war allerdings organisatorisch und finanziell zu schwach ausgestattet, um auf Dauer erfolgreich arbeiten zu können, so dass mit Unterstützung mehrerer Ministerien Preußens und des Reichs Mitte der 1920er Jahre eine grundlegende Neuorganisation stattfand: Es begann die Geschichte der *Jahresberichte für deutsche Geschichte* im engeren Sinne<sup>6</sup>, die als eine Art nationales Pendant zur zeitgleichen Initiative für die Gründung der *Internationalen Bibliographie der Geschichtswissenschaften* konzipiert waren. Als Herausgeber fungierten Albert Brackmann und Fritz Hartung; in die Trägerorganisation waren die Direktoren der Preußischen Staatsarchive und der Preußischen Staatsbibliothek sowie die Deutsche Bücherei Leipzig eingebunden.<sup>7</sup> Bereits für den ersten Band konnten sich Brackmann und Hartung auf eine breite Basis von 74 Berichterstellern stützen.<sup>8</sup> Rasch wurden die *Jahresberichte für deutsche Geschichte* zu einem unverzichtbaren Bestandteil der Fachinformation – die ab 1928 erarbeitete Neuauflage des *Dahlmann-Waitz* stützte sich nicht nur in inhaltlicher Hinsicht maßgeblich auf die *Jahresberichte*, sondern wurde auch im Auftrag der Gesellschaft *Jahresberichte für deutsche Geschichte* herausgegeben<sup>9</sup>, die sich damit gleichsam anschickte, zu einer zentralen Trägerstruktur für die geschichtswissenschaftliche Fachinformation in Deutschland zu werden.

Vom konzeptionellen Aufbau der Bände her wiesen die *Jahresberichte* der Zwischenkriegszeit gegenüber ihren Vorläufern eine wesentliche Änderung auf: Waren die bibliografischen Nachweise zuvor lediglich in den Fußnoten zu den Forschungsberichten enthalten, so wurden sie nun

5 Vgl. Loewe, Victor; Stimming, Manfred (Hgg.), *Jahresberichte der deutschen Geschichte*, 7 Bände, Berichtsjahre 1918-1924, Breslau 1920-1926.

6 Vgl. Brackmann, Albert; Hartung, Fritz (Hgg.), *Jahresberichte für deutsche Geschichte*, 16 Bände, Berichtsjahre 1925-1940, Leipzig 1927-1942.

7 Vgl. Brackmann u.a. (Hgg.), *Jahresberichte*, 3 (1927), Leipzig 1929 (wie Anm. 6), S. Vf. Vgl. auch Dietrich, Dieter, *Der Beitrag der Deutschen Bücherei zur Erarbeitung historischer Fachbibliographien seit 1929*, in: *Jahrbuch der Deutschen Bücherei* 25 (1989), S. 50-68.

8 Vgl. Brackmann u.a. (Hgg.), *Jahresberichte*, 1 (1925), Leipzig 1927 (wie Anm. 6), S. XIIIff.

9 Vgl. Haering, Hermann (Hg.), *Dahlmann; Waitz* (Hgg.), *Quellenkunde der Deutschen Geschichte*, 9. Auflage, Leipzig 1931, S. IIIff. u. VI.

von diesen separiert und traten, zusammengefasst in bibliografische Abschnitte, gleichberechtigt neben diese. Mit den entsprechenden Querverweisungen entstand somit ein Informationsmittel, dessen komplexe Struktur auch für die Digitalisierung besondere Herausforderungen und Möglichkeiten mit sich brachte.

In inhaltlicher Hinsicht besitzen die *Jahresberichte* der Zwischenkriegszeit als Nachweis der damaligen Fachliteratur in den bibliografischen Teilen, samt zeitgenössischer Einordnung in den Forschungskontext in den Berichtsteilen, nach wie vor einen ungebrochenen Wert: Sie bieten in ganz besonderer Weise einen Zugang zur historischen Forschung in der Weimarer Republik und der NS-Zeit, der für zahllose Fragestellungen gewinnbringend genutzt werden kann. Über die aus den Literaturzitate und Forschungsberichten implizit und explizit sichtbar werdenden Tendenzen und Wertungen hinaus sind die *Jahresberichte* aber auch mit ihrer inhaltlichen Gliederung und den Vorworten ein Spiegel dafür, wie politische und ideologische Motivationen die Historiografie beeinflussten: Die Erfahrung der Niederlage im Ersten Weltkrieg und der Gebietsabtretungen infolge des Versailler Vertrags fand nicht zuletzt darin ihren Widerhall, dass die *Jahresberichte*, „wie es sich von selbst versteht, von dem historischen Deutschland, nicht vom heutigen Umfang des Deutschen Reiches“ ausgingen. Sie „berücksichtigen daher auch die Gebiete, die einst zum Deutschen Reiche gehört haben, aber im Laufe der Geschichte von ihm getrennt worden sind. Wir behandeln ferner diejenigen Glieder des deutschen Volkes, die außerhalb des politischen Reichsverbandes stehen, bringen also einen Abschnitt über das Deutschtum im Auslande, der in den alten Jahresberichten fehlte, und ziehen auch die ausländische Literatur zur deutschen Geschichte in weitestem Umfang heran ...“<sup>10</sup> Die auch in der Geschichtswissenschaft zunehmenden völkischen Tendenzen ließen die *Jahresberichte* nicht unberührt: Die Herausgeber sahen sich 1931 mit der Aufforderung konfrontiert, Forschungen zur „Geschichte der deutschen Bevölkerung, die nach sozialen, statistischen, rasse- und siedlungsgeschichtlichen Gesichtspunkten untersucht werden solle“, in einem eigenen

10 Brackmann u.a. (Hgg.), *Jahresberichte*, 1 (1925) Leipzig 1927 (wie Anm. 6), S. V.

Bericht zu behandeln, was jedoch „aus praktischen wie auch aus sachlichen Gründen“ zunächst noch abgelehnt wurde.<sup>11</sup> Demgegenüber tritt im ersten nach der nationalsozialistischen Machtübernahme erschienenen Band deutlich das Bemühen hervor, die bisherige Arbeitsweise als von vornherein konform mit den neuen politischen Gegebenheiten darzustellen. Der „epochemachenden Bedeutung des Jahres 1933“ wurde durch eine Erweiterung des bislang mit dem Ersten Weltkrieg endenden Berichtgegenstandes um die „nunmehr abgeschlossene Periode von 1919 bis 1933“ Rechnung getragen.<sup>12</sup> In den Folgejahren wurde die inhaltliche Gliederung sukzessive um Abschnitte wie „Rasse und Bevölkerung“, „Die Juden in Deutschland“ oder „Staatsanschauungen und völkischer Gedanke“ erweitert; der Bereich „Bevölkerungsgeschichte“ erfuhr ab 1936 eine weitere Auffächerung.<sup>13</sup>

Die *Jahresberichte* der Zwischenkriegszeit können somit als Quelle von weitreichender Relevanz gelten. Um sie für die Forschung besser nutzbar zu machen, wurden die Bände in einem DFG-geförderten Kooperationsprojekt zusammen mit dem Trierer Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften bis 2003 vollständig digitalisiert. Ziel war es dabei einerseits, die Bände unter weitgehender Beibehaltung ihrer originalen Struktur in das elektronische Medium zu überführen und zugleich dessen technische Möglichkeiten für verschiedene Recherchezugriffe nutzbar zu machen. Das Projektergebnis konnte 2003 in einer ersten Version zur Verfügung gestellt werden. Nach Projektende wurde das Informationsangebot in einem weiteren Arbeitsschritt an der Berlin-Brandenburgischen Akademie der Wissenschaften überarbeitet und im März 2005 als „TELOTA-Projekt des Monats“ vorgestellt.<sup>14</sup> Kern der Maßnahme war eine Migrierung der ursprünglich SGML-codierten Daten in eine native XML-Datenbank.

11 Siehe Brackmann u.a. (Hgg.), *Jahresberichte*, 5 (1929), Leipzig 1931 (wie Anm. 6), S. Vf.

12 Siehe Brackmann u.a. (Hgg.), *Jahresberichte*, 7 (1931), Leipzig 1934 (wie Anm. 6), S. Vf.

13 Siehe Brackmann u.a. (Hgg.), *Jahresberichte*, 8 (1931), Leipzig 1934; 9/10 (1933/1934), Leipzig 1936; 12 (1936), Leipzig 1937 (alle wie Anm. 6).

14 Siehe <[http://www.bbaw.de/pom/langtext\\_200503.html](http://www.bbaw.de/pom/langtext_200503.html)> (25.10.2006).

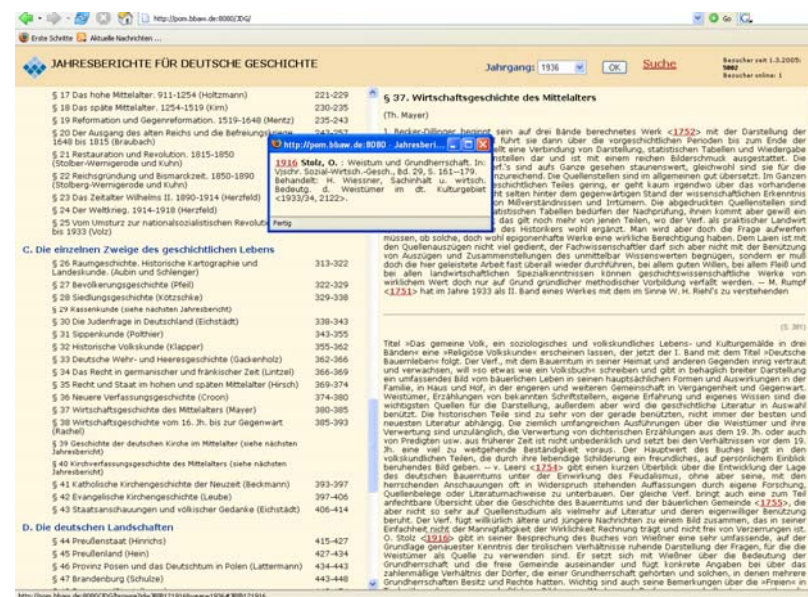


Abbildung 1: Blättern Zugriff auf Forschungsberichte, Verknüpfungen zu bibliografischen Angaben

Die digitalen *Jahresberichte für deutsche Geschichte aus der Zwischenkriegszeit* (Bände 1-14, Berichtsjahre 1925-1938)<sup>15</sup> bieten zunächst einen blättern Zugriff über die einzelnen Bände, wobei Inhaltsverzeichnis und Register weitestgehend in der Originalform beibehalten wurden (siehe Abbildung 1). Einen erheblichen Mehrwert gegenüber der Druckausgabe stellen die bandinternen Verknüpfungen dar, die nicht nur eine schnelle Navigation ermöglichen, sondern darüber hinaus in den bibliografischen Teilen sichtbar machen, ob ein Titel in den Forschungsberichten Berücksichtigung fand (siehe Abbildung 2). Durch die Implementierung bandübergreifender Suchmöglichkeiten ist es möglich, sich unter anderem einen Überblick über die Entwicklung bestimmter Forschungsgegenstände zu verschaffen (siehe Abbildung 3).

15 Siehe <<http://pom.bbaw.de:8080/JDG/>> (25.10.2006).

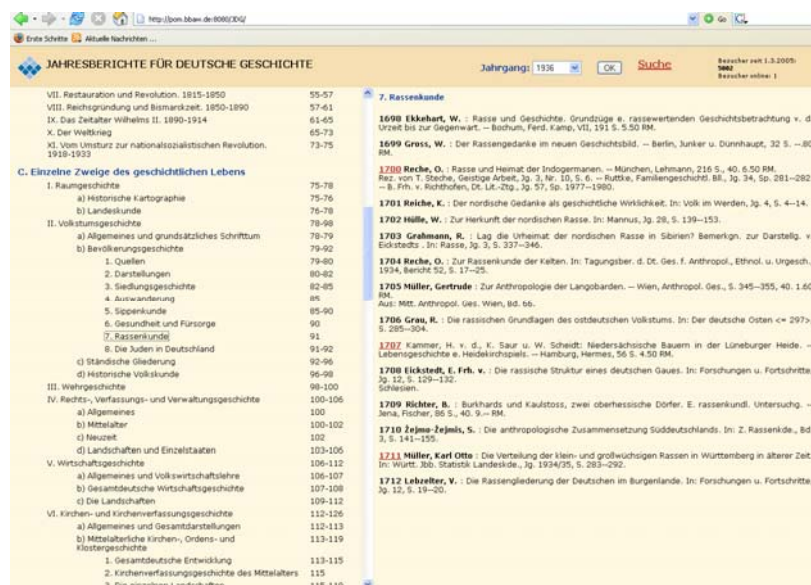


Abbildung 2: Bibliografischer Abschnitt mit Verknüpfungen zu Forschungsberichten

Die Arbeiten an dem Projekt sind noch nicht abgeschlossen: Redaktionelle Schwierigkeiten, die aufgrund der Komplexität des Informationsangebots, der großen Mitarbeiterzahl etc. schon bei der Erstellung der Druckbände vorhanden waren, machen sich noch heute bei der Überführung in die digitale Form bemerkbar. Die automatische Erkennung und Codierung uneinheitlich gestalteter bzw. teilweise auch fehlerhafter Bezüge wirft vielfach Probleme auf, so dass ein hoher Aufwand für die Nachbearbeitung erforderlich ist. Als weitere Arbeitsschritte sind neben der Überarbeitung die Einrichtung einer XQuery-Schnittstelle für individuelle Recherchefragen sowie die Konvertierung der bibliografischen Daten geplant, um diese, zusätzlich zu dem in sich geschlossenen Angebot, auch innerhalb der laufenden Datenbank der *Jahresberichte für deutsche Geschichte* bereitstellen zu können.

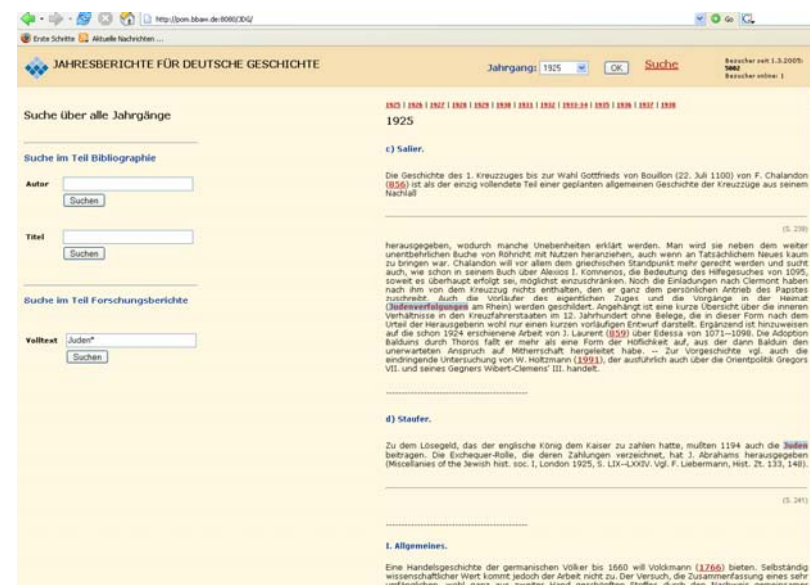


Abbildung 3: Bandübergreifende Recherchemöglichkeiten

\*\*\*

*Dr. Daniel Schlögl ist Arbeitsstellenleiter des Akademienvorhabens Jahresberichte für deutsche Geschichte an der Berlin-Brandenburgischen Akademie der Wissenschaften. E-Mail: schloegl@bbaw.de*