



Staatsarchiv der
Freien und Hansestadt Hamburg

Forschung in der digitalen Welt

Herausgegeben von Rainer
Hering, Jürgen Sarnowsky,
Christoph Schäfer und Udo
Schäfer

Veröffentlichungen aus dem Staatsarchiv der
Freien und Hansestadt Hamburg Band 20

Hamburg University Press

Forschung in der digitalen Welt
Sicherung, Erschließung und Aufbereitung von Wissensbeständen

Herausgegeben von Rainer Hering,
Jürgen Sarnowsky, Christoph Schäfer
und Udo Schäfer

Veröffentlichungen aus dem Staatsarchiv der
Freien und Hansestadt Hamburg

Band 20

Forschung in der digitalen Welt

Sicherung, Erschließung und
Aufbereitung von Wissensbeständen

Tagung des Staatsarchivs Hamburg und des Zentrums
,Geisteswissenschaften in der digitalen Welt'
an der Universität Hamburg am
10. und 11. April 2006

Herausgegeben von
Rainer Hering,
Jürgen Sarnowsky,
Christoph Schäfer und
Udo Schäfer

Hamburg University Press
Verlag der Staats- und Universitätsbibliothek
Hamburg

Impressum

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Die Online-Version dieser Publikation ist auf der Verlagswebsite frei verfügbar (*open access*). Die Deutsche Nationalbibliothek hat die Netzpublikation archiviert. Diese ist dauerhaft auf dem Archivserver der Deutschen Nationalbibliothek verfügbar.

Open access über die folgenden Webseiten:

Hamburg University Press – <http://hup.sub.uni-hamburg.de>

Archivserver der Deutschen Nationalbibliothek – <http://deposit.d-nb.de>

ISBN-10 3-937816-27-5

ISBN-13 978-3-937816-27-2

ISSN 0436-6638

© 2006 Hamburg University Press, Hamburg

Rechtsträger: Staats- und Universitätsbibliothek Hamburg, Deutschland

Produktion: Elbe-Werkstätten GmbH, Hamburg, Deutschland

<http://www.ew-gmbh.de>

Bildnachweis: Der Abdruck aller Abbildungen erfolgt mit freundlicher Genehmigung der Autoren bzw. des Autors des jeweiligen Beitrags.

Inhaltsübersicht

Einleitung	7
<i>Die Herausgeber</i>	
Grußwort	11
<i>Karin von Welck</i>	
„Wie ist es eigentlich gewesen, wenn das Gedächtnis virtuell wird?“	13
Die historischen Fächer und die digitalen Informationssysteme	
<i>Manfred Thaller</i>	
Datenstandards in der Erschließung historischer Dokumente	29
<i>Patrick Sahle</i>	
Fachspezifische Indexierung von historischen Dokumenten I	43
Quellen zwischen Zeichenketten und Information – Beispiel Urkunden	
<i>Georg Vogeler</i>	
Fachspezifische Indexierung von historischen Dokumenten II	59
Ein Framework zur approximativen Indexierung semistrukturierter Dokumente	
<i>Markus Heller</i>	
Digitale Erschließung und Sicherung von aktuellen archäologischen Befunden	85
<i>Christoph Schäfer</i>	
Digitale Urkundenbücher zur mittelalterlichen Geschichte	93
<i>Jürgen Sarnowsky</i>	
Verborgен, vergessen, verloren?	109
Perspektiven der Quellenerschließung durch die digitalen <i>Regesta Imperii</i>	
<i>Dieter Rübsamen und Andreas Kuczera</i>	

Virtuelle Zusammenführung und inhaltlich-statistische Analyse der überlieferten Reichskammergerichtsprozesse	125
<i>Bernd Schildt</i>	
Konzepte zur Bereitstellung digitalisierter frühneuzeitlicher Quellen ...	143
<i>Thomas Stäcker</i>	
Archive in der digitalen Welt	153
Informationstransfer zwischen Verwaltung und Wissenschaft	
<i>Rainer Hering</i>	
Nutzung von Digitalisaten am Beispiel des Geheimen Staatsarchivs Preußischer Kulturbesitz	161
<i>Dieter Heckmann</i>	
Das Angebot der Archive in der digitalen Welt	169
Retrokonversion, Datenaustausch und Archivportale	
<i>Frank M. Bischoff und Udo Schäfer</i>	
Geschichtswissenschaft auf dem Weg zur E-History?	183
<i>Angeblika Schaser</i>	
Beitragende	189

Einleitung

Längst wird in den Geisteswissenschaften eine Vielzahl von Projekten betrieben, die sich mit Digitalisierung im weitesten Sinne befassen. Heute geht es nicht mehr darum, eine einzelne Datenbank ‚im stillen Kämmerlein‘ aufzubauen, größere, überregionale und über das World Wide Web abfragbare Lösungen werden gesucht und zum Teil schon realisiert. Mehr denn je vermisst man jedoch eine Standardisierung, die den dringend notwendigen Austausch erleichtern und die gemeinsame Nutzung strukturierter Daten ermöglichen oder vorantreiben kann. Dazu bedarf es des Austauschs und gemeinsamer Arbeitsansätze.

Vor diesem Hintergrund fand am 10. und 11. April 2006 am Staatsarchiv der Freien und Hansestadt Hamburg eine gemeinsam mit dem Zentrum ‚Geisteswissenschaften in der digitalen Welt‘ der Universität Hamburg veranstaltete Tagung statt, deren Hauptziel darin bestand, zumindest in einigen Bereichen eine Annäherung im Sinne kompatibler Strukturen und Denksätze zu erreichen. Grundlage dafür war der intensive, interdisziplinäre Austausch, der in zweifacher Hinsicht übergreifend geschah, und zwar

- *inhaltlich*: über die Epochen hinweg, und
- *formal*: in Bezug auf verschiedene Überlieferungsträger (Wissenschaft, Archiv und Verwaltung).

Bewusst wurden damit diejenigen, die Wissensarchive aufbauen und pflegen, mit denen, die diese im Rahmen ihrer Forschungen analysieren, in einem intensiven Dialog zusammengeführt. Das erschien insofern besonders wichtig, als ein derartiger Schritt bei den bisherigen Diskussionen um die Forschung und die Sicherung von Wissen in einer immer stärker vernetzten und digitalisierten Welt über die formalen Grenzen hinweg so noch nicht vollzogen wurde. So konnten gemeinsam, aus der Sicht beider ‚Seiten‘, Probleme aufgezeigt und Perspektiven entwickelt werden.

Dabei stand im Rahmen der Tagung gerade nicht das Generieren technischer Einzellösungen im Vordergrund, sondern vielmehr die Entwicklung langfristiger Konzepte für die Aufbereitung und Sicherung von Wissensbeständen unterschiedlicher Couleur mit digitalen Methoden. Insbesondere der interdisziplinäre Ansatz und die Kooperation von Wissenschaft und Archiv erbrachten weiter führende Ergebnisse und werden dazu beitragen, dass ‚das Rad nicht immer wieder neu erfunden‘ werden muss. So konnten Desiderate und Konzepte jenseits der rein technischen und damit sehr stark zeitgebundenen Lösungsmodelle ausgelotet werden.

Bewusst wurde daher auf interdisziplinäre Kooperation gesetzt. So waren die Disziplinen

- Archivwissenschaft,
- Informatik,
- Archäologie,
- Geschichtswissenschaften und
- Philologie

auf die eine oder andere Weise in das Konzept eingebunden. Bei den Überlegungen zur Nutzung digitaler Methoden ging es jeweils auch um reflektierte Langzeitlösungen. Der vorliegende Band legt – bewusst zeitnah – die Beiträge der Tagung im Druck vor, damit die Diskussion der Probleme auf neuer Grundlage weitergehen kann.

Der erste Teil der Beiträge ist Grundsatzfragen und grundlegenden methodischen Ansätzen gewidmet. Manfred Thaller setzt sich mit den Konsequenzen der Digitalisierung für die Geschichtswissenschaft auseinander, Patrick Sahle analysiert Modelle für die digitale Erfassung und Erschließung historischer Dokumente, und Markus Heller und Georg Vogeler stellen sowohl theoretische wie konkrete Konzepte für die Indexierung von Dokumenten vor. Im Anschluss daran werden im zweiten Teil des Bandes Projekte und Problemlösungen aus den verschiedenen Epochen der Geschichtswissenschaft diskutiert. Christoph Schäfer behandelt am Beispiel der Alten Geschichte Modelle für Wissenstransfer durch Visualisierung, Jürgen Sarnowsky, Andreas Kuczera und Dieter Rübsamen präsentieren ausgewählte Urkundenprojekte zur mittelalterlichen Geschichte, Bernd Schildt und Thomas Stäcker führen anhand frühneuzeitlicher Quellen in Konzepte datenbankgestützter Grundlagenforschung sowie in Digitalisie-

rungsprojekte ein. Der dritte Teil der Beiträge gilt schließlich den Chancen und Problemen des elektronischen Zugriffs auf Quellen- bzw. Archivbestände aus archivarischer Sicht. Rainer Hering arbeitet die Probleme digitaler Überlieferung der Behörden heraus, Dieter Heckmann macht an einem konkreten Beispiel Stand und Perspektiven der Digitalisierung von Archivmaterialien deutlich, und Frank M. Bischoff und Udo Schäfer stellen die Planungen für ein deutsches Archivportal vor. Den Abschluss bildet die Zusammenfassung von Angelika Schaser, die weiter führende Fragestellungen entwickelt.

Die Herausgeber danken der Thyssen-Stiftung für die großzügige Förderung der Tagung sowie Hamburg University Press für die vorzügliche Betreuung des Bandes.

Die Herausgeber

Grußwort

Karin von Welck

Sehr geehrte Damen und Herren,
sehr geehrter Herr Professor Thaller,

ich freue mich sehr, Sie alle im Staatsarchiv der Freien und Hansestadt Hamburg zu der Tagung *Forschung in der digitalen Welt* begrüßen zu dürfen. Vor zwei Jahren fand in diesem Raum, dem Lorichssaal, die Tagung *Digitales Verwalten – Digitales Archivieren* statt, deren Ergebnisse in den Veröffentlichungen aus dem Staatsarchiv der Freien und Hansestadt Hamburg publiziert worden sind. Damals ging es um die Schnittstelle zwischen Verwaltung und Archiv. Heute und morgen wird es auch um die Schnittstelle zwischen Archiv und Forschung, vor allem aber um die Nutzung von Datenbanken und Internet durch die historische Forschung gehen. Für die teilnehmenden Archivarinnen und Archivare wird – denke ich jedenfalls – von besonderem Interesse sein, wie die moderne Forschung mit den sich bietenden technischen Möglichkeiten arbeitet und welche Erwartungen der Forschung an die Archive sich aus diesen Arbeitsformen heraus entwickeln oder entwickeln können. Der für die Tagung gewählte interdisziplinäre Ansatz wird sich deshalb – davon bin ich überzeugt – als besonders fruchtbar erweisen.

Man könnte denken, das ‚Digitale‘ würde nun zum gemeinsamen Nenner der Tagungen, die das Staatsarchiv durchführt. Es war jedoch eher der interdisziplinäre Ansatz, der das Staatsarchiv veranlasste, gemeinsam mit dem Zentrum ‚Geisteswissenschaften in der digitalen Welt‘ diese Veranstaltung zu planen. Da Herr Professor Sarnowsky bereits seit längerem zusammen mit dem Staatsarchiv an dem Projekt eines *Virtuellen Hamburgischen Urkundenbuchs* arbeitet, bot sich diese Kooperation geradezu an. Im Rahmen seiner Möglichkeiten wird das Staatsarchiv den Studierenden der vom Zentrum angebotenen Studiengänge auch Praktika ermöglichen. Dieses

Engagement für die Ausbildung freut mich natürlich ganz besonders. Das Staatsarchiv wird den interdisziplinären Ansatz auch in Zukunft weiter verfolgen. Als Amt der Kulturbehörde hat es ein sehr breites Aufgabenspektrum, so dass sehr unterschiedliche Partner denkbar sind: Nicht nur Museen und Bibliotheken, nicht nur die historischen Wissenschaften, sondern auch die Rechtswissenschaft oder – man denke an die vielen in diesem Haus verwahrten Karten und Pläne – die Geowissenschaften kommen hier in Betracht. Ich begrüße es sehr, dass das Staatsarchiv von Zeit zu Zeit versuchen wird, in Kooperation mit anderen Einrichtungen Gastgeber eines interdisziplinären Austauschs zu sein.

„Digitale Bibliotheken“, „digitale Archive“: Diese Themen sind mittlerweile auch in der Politik angekommen. Im letzten Jahr hat die Europäische Kommission dem Rat und dem Europäischen Parlament eine Mitteilung vorgelegt, in der sie sich intensiv mit der Online-Verfügbarkeit des europäischen Kulturerbes beschäftigt, wobei ich sehr froh bin, dass die Kommission ihr Augenmerk nicht nur auf das Thema dieser Tagung, nämlich die Zugänglichkeit, richtet, sondern sich auch mit der langfristigen Bewahrung des digitalen Kulturerbes befasst. Mit dieser Mitteilung und anderen sich auf die Digitalisierung von Kulturgut beziehenden Themen hat sich auch der Kulturausschuss der Kultusministerkonferenz in seiner letzten Sitzung vor wenigen Wochen beschäftigt. Sie sehen, das Thema gewinnt auch auf der politischen Ebene an Fahrt. Die Frage nach einer nationalen Digitalisierungsstrategie steht auf der Tagesordnung. Allerdings ist deren Verwirklichung nicht ohne erhebliche finanzielle Mittel denkbar. Als Mitglied einer Landesregierung weiß ich nur zu gut um die Situation der öffentlichen Haushalte Bescheid. Ob alternative Wege zur Finanzierung möglich sind, bleibt abzuwarten. Eines muss aber auf jeden Fall erreicht werden: die enge Kooperation zwischen Bund und Ländern, zwischen Bibliotheken, Archiven, Museen und anderen Kulturgut sichernden Einrichtungen sowie mit der Deutschen Forschungsgemeinschaft und anderen die Wissenschaft fördernden Gesellschaften. In unserem föderalen System ist dieses keine leichte, aber eine lohnende, die kulturelle Vielfalt auch in der digitalen Welt bewahrende Aufgabe.

Ich wünsche Ihnen allen eine interessante und ertragreiche Tagung. Soweit Sie von auswärts kommen, bleibt Ihnen vielleicht am Dienstag noch ein wenig Zeit, unser wunderschönes Hamburg zu genießen oder wenigstens in der Nähe des Hauptbahnhofs einen Blick in die Kunsthalle oder das Museum für Kunst und Gewerbe zu werfen.

„Wie ist es eigentlich gewesen, wenn das Gedächtnis virtuell wird?“

Die historischen Fächer und die digitalen Informationssysteme

Manfred Thaller

1. Umfeld

Ranke's scheinbar bescheidener Satz „Man hat der Historie das Amt, die Vergangenheit zu richten, die Mitwelt zum Nutzen zukünftiger Jahre zu belehren, beigemessen: so hoher Ämter unterwindet sich gegenwärtiger Versuch nicht: er will bloß zeigen wie es eigentlich gewesen.“¹ ist wohl auch heute noch ein guter Ausgangspunkt für die Bewertung der Relevanz neuerer Entwicklungen für die historischen Disziplinen. Nicht weil man ihn als Ausdruck naiven Objektivitätsglaubens unkritisch übernehmen muss; gerade für die Kritik des sich in ihm darstellenden Objektivitätsglaubens stellt er einen vorzüglichen Ausgangspunkt dar. Aber mindestens deshalb, weil er in seiner scheinbaren Bescheidenheit doch Ausdruck einer sich selbst als geschlossen und konsistent sehenden Historischen Wissenschaft ist, was man der heutigen defensiven Debatte um die anhaltende oder sich verlierende Bedeutung der historischen Disziplinen nur wünschen kann. Ein guter Ausgangspunkt ist der Satz gerade auch deshalb, weil er in anderen Wissenschaftslandschaften noch immer als spezifisch deutsch gilt – so, wenn in den 90er Jahren ein amerikanischer Versuch, den

¹ Leopold von Ranke: Vorwort zur ersten Ausgabe (1924) der Geschichte der Romanischen und Germanischen Völker. Leipzig ³1885. VII.

deutschen Stand des Rechnereinsatzes in den historischen Disziplinen unter dem Titel *Counting ‚How it Really Was‘*² veröffentlicht wurde.

Das Umfeld historischer Forschung ändert sich ständig; eine der Änderungen, denen es eben jetzt, im Jahre 2006, unterliegt, besteht darin, dass der ‚Rohstoff‘ historischer Forschung, als die ungedruckten und gedruckten Quellen, unverhältnismäßig leichter zugänglich werden könnte, als er dies bisher war, wenn man einige vollkommen unabhängig von den historischen Disziplinen entstandenen Trends fortschreibt. Wir sprechen dabei von der systematischen Bereitstellung von Digitalisaten von Bibliotheks- und Archivbeständen, die national und international innerhalb zu einem Zeitpunkt vorangetrieben wird, zu dem die Vorstellung, dass die neuen Möglichkeiten des digitalen Informationsaustausches und der Vernetzung von Information, den für ihre Analyse benötigten Werkzeugen zusammen mit neuen Publikationswegen eine ganz allgemein ‚verbesserte‘ Art der Forschung ermöglichen, was sich hinter dem forschungspolitisch wichtigen Schlagwort der E-Science verbirgt (wobei ‚E‘ für *enhanced* steht)³. Aus dieser Vorstellung entwickelt sich in anderen Ländern mittlerweile ein Konzept der ‚E-Humanities‘, rund um die Vorstellung einer Geisteswissenschaft, die die digitalen Medien langfristig als primäres Mittel für den Zugang zu Forschungsinformation, deren Analyse und der Veröffentlichung der so erzielten Ergebnisse nutzen könnte. Dies kann, recht stark verengt, durch einen Schwerpunkt auf der Untersuchung der Konsequenzen durch ein Forschungsprogramm geschehen (wie in den Niederlanden)⁴, durch die weitere Neufokussierung des seit mehreren Jahren betriebenen Neuaufbaus einer IT Infrastruktur für die Geisteswissenschaften (wie in Großbritannien)⁵ oder durch nationale Vorstudien für eine ‚Cyberinfrastruktur‘ für die Geisteswissenschaften (wie derzeit in den USA)⁶.

² E. A. Johnson: *Counting ‚How It Really Was‘: Quantitative History in West Germany*. In: *Historical Methods Newsletter* 21 (1988). S. 63.

³ Vgl. *Wissenschaftsmanagement – Zeitschrift für Innovation, Special 1/2005* „Management im virtuellen Forschungsraum“.

⁴ Adresse: <http://www.virtualknowledgestudio.nl> (letzte Einsichtnahme am 19.06.2006; zu diesem Datum wurde auch die Sichtbarkeit aller weiteren angeführten URLs geprüft).

⁵ Vgl. <http://www.ahrb.ac.uk/>; eine absehbar langfristig verfügbare Beschreibung der derzeitigen Programme kann nicht angegeben werden.

⁶ Vgl. <http://www.acls.org/cyberinfrastructure/cyber.htm>.

Konkreter als diese weitgreifenden Planungen, die vielleicht so weit greifen, dass sie niemals konkrete Formen annehmen ist, dass in den letzten Jahren erhebliche Mengen von digitalisierten Bibliotheks- und Archivmaterialien verfügbar gemacht worden sind – in Deutschland etwa durch ein einschlägiges Förderprogramm der DFG, das in einer ersten Förderphase bis 2003 etwa 90 Projekte gefördert hat, die Quellen eines sehr breiten Spektrums, von antiken Papyri bis zu Materialien zur Geschichte der DDR, bereitgestellt haben.⁷ Trotz mancher Probleme im Detail waren die Erfahrungen damit hinreichend überzeugend, dass die weitere Förderung so ausgerichtet werden wird, dass von der Digitalisierung einzelner Objekte zur Digitalisierung ganzer Bestände übergegangen wird⁸, wobei das strategische Ziel, sämtliche Drucke des 16. und 17. Jahrhunderts, die in den deutschen nationalbibliographischen Verzeichnissen deutscher Drucke zu diesen Jahrhunderten enthalten sind, zu digitalisieren, mittlerweile durch eine erste Ausschreibung für einschlägige Projekte konkretisiert wurde.

Andere Erfahrungen im Bereich der systematischer Digitalisierung führten schließlich Ende 2004 zur Absichtserklärung von Google, in Zusammenarbeit mit einigen großen Bibliotheken systematisch Druckwerke im Umfang von einigen Millionen Bänden digital bereit zu stellen.⁹ Dieses Projekt als solches ist überschattet von einer gewissen Unklarheit der tatsächlichen Ziele; für den europäischen Bereich wichtiger ist wohl, dass diese Ankündigung vor allem – aber nicht nur – durch den Direktor der Bibliothèque National de France, Jean-Noël Jeanneney, als Herausforderung an die kulturelle Präsenz Europas im Internet dargestellt wurde.¹⁰ Während Bestrebungen zur systematischen Digitalisierung des kulturellen Erbes bisher vor allem in Kreisen von Bibliothekaren und Archivaren diskutiert wurden, genießen sie seit kurzem hohe politische Aufmerksamkeit – und die europäische Kommission hat angekündigt in den nächsten fünf Jahren ab 2006 mindestens fünf Millionen „Bücher und andere Dokumente“ öffentlich bereitzustellen.

⁷ Vgl. den vom Verfasser redigierten Evaluierungsbericht über dieses Förderprogramm: http://www.dfg.de/forschungsforderung/wissenschaftliche_infrastruktur/lis/download/retro_digitalisierung_eval_050406.pdf.

⁸ Vgl. das einschlägige Strategiepapier der DFG unter http://www.dfg.de/forschungsforderung/wissenschaftliche_infrastruktur/lis/download/pos_papier_kulturelle_ueberlieferung_0511.pdf.

⁹ Vgl. <http://books.google.com/> zum aktuellen Angebot.

¹⁰ Jean-Noël Jeanneney: Googles Herausforderung. Für eine europäische Bibliothek. Berlin 2006.

Kleinere Beiträge zu diesem Bemühen um die systematische Bereitstellung digitalisierter Quelleninformationen stammen auch vom Verfasser: Beginnend mit der Bereitstellung eines geschlossenen Bestandes aus dem Stadtarchiv Duderstadt in den 90er Jahren¹¹, fortgesetzt durch die digitale Umwandlung der Codices der Kölner Diözese und Dombibliothek¹² und zuletzt weitergetrieben durch ein Projekt zur systematischen Digitalisierung von Inkunabeln¹³ einerseits, das nachweisen sollte, dass eine gezielte, flächendeckende Digitalisierung der deutschen Inkunabeln realisierbar wäre, andererseits durch eine größere Anzahl von Kleinprojekten, die zeigen sollten, dass derartige Digitalisierungsprojekte bereits mit geringen Mitteln durchgeführt werden können.¹⁴ Beschreibungen dieser Projekte sind in letzter Zeit mehrfach im Druck erschienen¹⁵; sie sollen hier daher nicht wiederholt werden.

2. Konsequenzen für die Historischen Fächer und die von ihnen genutzten Infrastrukturen¹⁶

Vor diesem Hintergrund ist gegenwärtig eine Art stiller Diskussion über die Frage, was den eigentlichen Nutzen dieser von mancher Seite lebhaft propagierten neuen digitalen Technologien innerhalb des Bereichs der Geistes- und Kulturwissenschaften ausmache, entstanden. Mit „stiller Dis-

¹¹ Manfred Thaller und Hans-Heinrich Ebeling: *Digitale Archive. Die Erschließung und Digitalisierung des Stadtarchivs Duderstadt*. Göttingen 1999 (vgl. <http://www.archive.geschichte.mpg.de/duderstadt/>).

¹² Adresse: <http://www.ceec.uni-koeln.de>.

¹³ Adresse: <http://inkunabeln.ub.uni-koeln.de>.

¹⁴ Vgl. dazu die Projektübersicht unter <http://www.hki.uni-koeln.de/projekte/projekte-b.html>.

¹⁵ Manfred Thaller: *Digital manuscripts as base line for digital editions*. In: *Digital Technology and Philological Disciplines*. Hg. von Andrea Bozzi, Laura Cignoni, Jean-Louis Lebrave (*Linguistica Computazionale* 20/21). Pisa, Rom 2004. S. 489–511. – Ders.: *Reproduktion, Erschließung, Edition, Interpretation: Ihre Beziehungen in einer digitalen Welt*. In: *Vom Nutzen des Edierens*. Hg. von Brigitte Merta, Andrea Sommerlechner und Herwig Weigl (*Mitteilungen des Instituts für Österreichische Geschichtsforschung Ergänzungsband* 47). Wien 2005. S. 205–227. – Ders.: *Alte Archive – Neue Technologien*. In: *Alte Archive – Neue Technologien*. Hg. von Thomas Aigner. Wien 2006 (im Druck).

¹⁶ Dieser und der folgende Abschnitt wiederholen Positionen, die bereits im in der vorigen Anmerkung zitierten Aufsatz des Verfassers „Reproduktion, Erschließung, Edition, Interpretation ...“ ausgearbeitet wurden. Die Darstellung wurde großteils unverändert übernommen.

kussion' meinen wir, dass sehr unterschiedliche Positionen existieren, die jedoch in der Regel nicht explizit gemacht, noch weniger verschriftlicht werden, sondern meist aus programmatischen Erklärungen zum Nutzen (oder der Bedrohlichkeit) der neuen Informationstechnologien für die Bereitstellung historischen Quellenmaterials erschlossen werden müssen.

Bevor wir solche Positionen einander gegenüberstellen, möchte ich zunächst herausstellen, was meiner Meinung nach die wirklich entscheidende Änderung bei den Voraussetzungen für die Publikation von Quellen in den letzten Jahren ist. Und auch wenn ich zögere, das Wort ‚Revolutionen‘ zu verwenden, gerade weil bestimmte Bereiche der Medienwissenschaft geradezu ständig damit beschäftigt sind, immer neue ‚Medienrevolutionen‘ zu entdecken, so glaube ich doch, dass es eine ganz bestimmte technologische Änderung in unseren Arbeitsbedingungen gibt, die langfristig extrem gravierende Auswirkungen haben wird. Dabei geht es um folgendes Phänomen:

Seit den frühesten Versuchen der Menschheit, Information zu bewahren und/oder sie zu vervielfältigen und zu vermitteln, galt eine unumstößliche Regel: Sogar die beste und sorgfältigste verbale Beschreibung eines Objekts war immer billiger als jede auch nur einigermaßen befriedigende visuelle Reproduktion. Dies ist ganz offensichtlich so für die Antike, in der wirklich detailgenaue Abbildungen kaum möglich waren, aber selbst nach der Erfindung der Photographie behielt diese Regel bis vor kurzem ihre Gültigkeit. Man vergleiche bis in die allerjüngste Zeit den Preis reich illustrierter, aber nicht weiter komplex strukturierter kunsthistorischer Bildbände und extrem sorgfältig redigierter, aber bildarmer Texteditionen. Aus dieser kulturellen/technologischen/ökonomischen Voraussetzung entstand eine Tradition des Umgangs mit Quellen, die die Annahme widerspiegelt, dass die textuelle Wiedergabe wesentlich leichter finanzierbar ist als die visuelle Reproduktion. Es galt immer die Regel:

- Fasse die weniger wichtigen Texte zusammen (Regesta).
- Veröffentliche die wichtigsten von ihnen kritisch (Editio critica).
- Reproduziere ausgewählte Einzelstücke visuell (Facsimile).

Mit modernen digitalen Technologien trifft, wie wir oben hergeleitet haben, die Grundannahme ‚Beschreibung ist billiger als Reproduktion‘ nicht mehr länger zu. Die visuelle Erscheinung eines Textes, oder von extrem großen

Textmengen, zu reproduzieren ist sehr viel billiger als diese Texte zu transkribieren, zusammenzufassen, zu edieren, mit anderen Worten: zu beschreiben. Sogar wenn wir „nur“ über Zusammenfassungen in der Tradition von Regesten sprechen, ist es ganz klar, dass kein Textcorpus für 50 Cent bis 1 Euro pro Ausgangsseite registrierbar ist.

Wenn diese Annahme nicht mehr stimmt, was für Konsequenzen hat dies für unser Bestreben, Dokumente aus der Vergangenheit besser zugänglich zu machen? Spontan können wir feststellen, dass es jederzeit möglich wäre, die für die gesamte historische Forschung zugänglichen Dokumente – die also für darüber stattfindende Debatten allen daran Teilnehmenden auch wirklich zugänglich sind – mit relativ geringen Mitteln extrem auszuweiten, wenn wir davon ausgehen, dass wir Dokumente, die wir sofort zu Gesicht bekommen, mit sehr viel spärlicheren Angaben beschreiben können als solche, bei denen die Güte der Beschreibung letzten Endes als Grundlage der Entscheidung dienen muss, ob sich die beträchtlichen Kosten des unter Umständen mit einer Reise verbundenen Zugriffs auf das Original lohnen. Wie wir diese Aussage bewerten, hängt letztlich von der Position ab, die der Einzelne in der oben erwähnten stillen Diskussion einnimmt.

Eine dieser Positionen kann etwa folgendermaßen explizit gemacht werden: Bloß weil sich die Technik der Publikation ändert, ändert sich noch lange nichts an der methodischen und intellektuellen Situation des Herausgebers. Moderne Technologien sollten deshalb genutzt werden, um den traditionellen Arbeitsprozess zu optimieren. Dies hat vier wichtige Konsequenzen:

- Die Edition in einer digitalen Umwelt sollte so nah wie möglich das gedruckte Paradigma wiedergeben. Dies impliziert auch, dass mehrere kritische Apparate am Bildschirm genauso zu reproduzieren sind wie auf einer gedruckten Seite.
- Was bisher am publikationswürdigsten war, bleibt es. Die bisher als wichtigste Quellen betrachteten Dokumente sollten auch die ersten sein, die mit den neuen Technologien publiziert werden. (optimierter Zugang zu einem kanonischen Corpus). Das wichtigste Dokument in der Geschichte eines Landes bleibt stets der erste Kandidat für die Anwendung neuer Reproduktionsmöglichkeiten: als Erstes im Druck ediert, als Erstes lithographisch wiedergeben, Pilotobjekt für die verschiedenen Generationen photographischer Technik. Konsequenterweise nimmt die Magna Carta eine prominente Stelle auf den Seiten der

British Library ein und die Gutenberg-Bibel ist bereits mehrfach, man ist versucht zu sagen: vielfach, digitalisiert worden.

- Ein Editor publiziert noch immer ein abgeschlossenes, bis in die letzte Einzelheit durchdachtes und auf sehr lange Zeit angelegtes Werk, für das er oder sie allein verantwortlich ist. Diese Verantwortung zwingt zu äußerster Sorgfalt: Eine Edition ist als Ganzes erst verfügbar, wenn der Editor überzeugt ist, dass sie so perfekt ist, wie sie dies nach menschlichem Vermögen sein kann.
- Bibliothekar und Archivar sind die Wächter ihrer Bestände. Sie kontrollieren den Zugang zu den Quellen. Nur wenn Material gemäß dem neusten Stand der relevanten Wissenschaft erschlossen wurde, kann es einem Benutzer oder einer Benutzerin anvertraut werden, sei es digital, sei es physisches Original.

Lassen Sie mich betonen, dass ich diese Meinung für in sich konsistent, logisch geschlossen und überaus ehrenwert und respektabel halte. Sie respektierend halte ich sie dennoch für falsch und plädiere dafür, unsere Vorstellung vom Zugang zu historischen Dokumenten weiter zu entwickeln. Viele der Details, die sich aus diesem Vorschlag ergeben, werden noch intensiver Klärung bedürfen. Trotzdem würde ich behaupten, dass die folgende Position in sich eben so konsistent ist wie die eben vorgetragene.

Das typographische Aussehen von Editionen ist nicht das Ergebnis eines bewussten teleologischen Prozesses: Es hat sich sukzessive aus dem Versuch ergeben, das Beste aus den Bedingungen und Grenzen der Drucktechnik zu machen. Genau wie die ersten gedruckten Bücher versuchten, wie Codices¹⁷ auszusehen, lag es nahe, die ersten digitalen Publikationen nach dem Beispiel gedruckter Bücher zu gestalten; genau wie gedruckte Medien ihre eigene Form schließlich gefunden haben, sollten die digitalen Medien ihre eigene suchen.¹⁸ Der grundlegende Zweck der modernen Technologien

¹⁷ Dass übrigens der Codex bereits selbst das Ergebnis einer Reihe von Konsequenzen „publikationstechnischer Erfindungen“ war, die das Verhältnis zum Text fundamental verändert haben, wird in Ivan Illichs viel zu wenig beachteter Studie zu Hugo von St. Victor sehr schön geschildert: Ivan Illich: *In the Vineyard of the Text. A Commentary to Hugh's Didascalicon*. Chicago, London 1993.

¹⁸ Vgl. als „Zustandsbericht“ Gesine Boesken: *Lesen am Computer – Mehrwert oder mehr Verwirrung? Untersuchungen zur „Konkurrenz“ zwischen Buch und Hypertext*. In: *Jahrbuch für Computerphilologie* 4 (2002). S. 85–114; auch unter <http://computerphilologie.uni-muenchen.de/jg02/boesken.html>.

innerhalb der editorischen Arbeit ist deshalb die Unterstützung eines weiter entwickelten Arbeitsprozesses. Dieser wird umso produktiver sein, wenn er die hervorstechenden Eigenschaften der neuen Technologien ausnützt. Dazu gehört das Prinzip des Verteilten Arbeitens bei dem Material, zu dessen Bereitstellung viele zusammenwirken, wie im Falle der Inkunabeln und dem digitalen Bildarchiv *prometheus*, wie ein integriertes Informationsangebot aussieht. Dazu gehört das Prinzip des kooperativen Arbeitens, das mehr Partner in den editorischen Prozess integriert als bisher.

Dies führt zu den folgenden vier Gegenpositionen zu den eben genannten:

- Die neuen Technologien sollten genutzt werden, um über das Bisherige hinausgehende, mit den bisherigen Techniken nicht gegebene Möglichkeiten zu schaffen.¹⁹ Wird ein editorisches Interesse am besten durch eine gedruckte Edition bedient, sollte man sie drucken.²⁰ Digitale Techniken sollten dem vorbehalten bleiben, was mit dem Druck nicht zu erreichen war. Beispielsweise sollte gezielt untersucht werden, ob es nicht Darstellungsweisen des Inhalts von kritischen Apparaten²¹ gibt, die sich auf inhärent dynamischen Medien wie dem Bildschirm besser darstellen lassen als im Druck.²² Wird die Verwandtschaft von zwanzig Zeugen in einem gedruckten Apparat wirklich transparent?
- Was die editorischen Wissenschaften in den letzten Jahrzehnten auf der Ebene des hohen von ihnen erreichten fachlichen Standards und unter hervorragender Ausnutzung der Drucktechnologie publiziert haben, ist für die Fachwelt verfügbar. Dieses Material mit Hilfe der neuen Technologien nochmals zu veröffentlichen, verbessert seine Zugänglichkeit bestenfalls marginal. Dem entspricht der Standpunkt, dass die neuen Technologien bevorzugt zur Veröffentlichung jener Typen von Material verwendet werden sollten, die entweder wegen spezieller Ei-

¹⁹ Vgl. Sören Steding: Warum noch Drucken? Über die Nachteile gedruckter Editionen. In: Jahrbuch für Computerphilologie 3 (2001). S. 149–159; auch unter <http://computerphilologie.uni-muenchen.de/jg01/steding.html>.

²⁰ Vgl. Dino Buzzetti: Diacritical ambiguity and Markup. In: Augmenting Comprehension (wie Anm. 7). S. 175–188, hier S. 175: „A digital edition can only be wholly justified if it is capable of providing viable solutions to problems that can poorly be faced in any other way“.

²¹ Oder, weniger oberflächlich, Konzepte von ‚Text‘, welche die durch Variantenapparate ausgedrückte Variabilität besser ausdrücken. Vgl. Dino Buzzetti: Rappresentazione digitale e modello del testo. In: Il ruolo del modello nella scienza e nel sapere (Contributi del Centro Linceo Interdisciplinare ‚Beniamino Segre‘ 100). Roma 1999. S. 127–161.

²² Vgl. Domenico Fiorimonte: Scrittura e filologia nell’era digitale. Torino 2003. S. 209–237.

genschaften oder wegen ihrer schieren Menge durch die alten Technologien nicht veröffentlicht werden konnten. (Dekanonisierung der Quellenkunde.) Sowohl die Magna Carta als auch die Gutenberg-Bibel sind hervorragende Beispiele für die Art von Material, die nach Ansicht des Verfassers nicht – und ganz bestimmt nicht vorrangig – in die digitale Form überführt werden sollten.

- Ein Herausgeber ist nach diesem Modell für eine wohl definierte Phase innerhalb eines länger andauernden Arbeitsprozesses verantwortlich, für den die wissenschaftliche Gemeinschaft als Ganzes verantwortlich ist. Das bedeutet, dass eine Edition als eine Serie aufeinander aufbauender Arbeitsgänge verstanden wird, die Schritt für Schritt Schichten zusätzlichen Wissens zur einfachen Reproduktion eines Stückes – oder einer Sammlung – hinzufügen. Die erste derartige Schicht besteht konzeptuell wirklich einfach in der raschen und kostengünstigen „nackten“ Digitalisierung – die keinerlei intellektuelles Problem löst, das Dokument aber aus seiner „Gefangenschaft“ im Archiv befreit und es der öffentlichen Diskussion zugänglich macht. Die erste Schicht könnte möglicherweise eine „nackte“ Digitalisierung sein, daran könnte sich als zweite Schicht ein sehr limitiertes Zugangssystem anschließen, das nicht mehr leistet als die Handhabung der Reproduktionen zu vereinfachen. Eine dritte Schicht könnte in einer akribischen, aber immer noch weitgehend mechanischen Transkription²³ bestehen, an die sich als vierte eine sorgfältige Kollation mit einem zweiten Textzeugen anschließt²⁴, die schließlich nach mehreren weiteren Schritten in eine kritische Edition im ursprünglichen Sinne mündet. Eine sechste kann ein Kommentar eines Sprachhistoriker zu der hier als fünfte Schicht angesprochenen Edition sein, eine siebente deren inhaltliche Interpretation, und so weiter.

²³ Der Vollständigkeit halber sei darauf hingewiesen, dass eine vom Menschen unbeaufsichtigte automatische Transkription von Manuskriptmaterial in absehbarer Zeit nach wie vor unwahrscheinlich ist. Für die technischen Ansätze dazu siehe Bianca Feldmann: OCR von Handschriften. In: *Codices Electronici Ecclesiae Coloniensis* (wie Anm. 5), S. 107–143. Als Erfahrungsbericht über an mittelalterlichem Material getestete Ansätze: *Computer-aided Recovery and Analysis of Damaged Text Documents*. Hg. von Andrea Bozzi. Bologna 2000.

²⁴ Zum Einsatz der Informationstechnologie bei der Kollationierung und davon ausgehend bei der Erstellung von *Stemmata* vgl. Peter Robinson: *Making Electronic Editions and the Fascination of What is Difficult*. In: *Digital Technology* (wie Anm. 2), S. 415–437.

- Archivar und Bibliothekar sind nach diesem Modell mehr als für alles andere dafür verantwortlich, das Material verfügbar zu machen. Kataloge und andere Werkzeuge der Erschließung der Bestände einer Einrichtung sind Werkzeuge zu einem bestimmten Ziel; sie sind niemals Selbstzweck. Nie und nimmer darf das Argument, dass ein Dokument nicht gut genug katalogisiert, verzeichnet, erschlossen sei, ein Grund dafür sein, es der Forschung nicht zur Verfügung zu stellen – und wenn es nur auf der Basis einer mechanischen Reproduktion ist.

Der für die meisten Leser dieser Sichtweise der Bedeutung der neuen Technologien strittigste Standpunkt ist natürlich der, dass Editionen als ein geschichteter Prozess miteinander kooperierender Forscher und nicht das monolithische Opus Magnum eines Einzelnen verstanden werden sollten. Lassen Sie mich deshalb dieses Argument noch ein wenig weiter verfolgen, um deutlicher zu erklären, wie ich mir die Implementation eines solchen „editorischen Prozesses“ vorstelle.

3. Konsequenzen für das Verhältnis zu Quellen und Editionen

„Die Edition“ einer Quelle kann gegenwärtig als eine Serie von Tätigkeiten angesehen werden, von denen jede abgeschlossen sein muss, bevor die nächste beginnt. Und jede von ihnen erzeugt Ergebnisse, die nicht mehr verändert werden können, sobald diese Tätigkeit beendet wurde. Dies führt zum folgenden Verständnis der Rollen²⁵ einzelner am Umgang mit historischen Quellen Beteiligten:

- (1) Archivar oder Bibliothekarin bewahren das Dokument.
- (2) Herausgeber oder Herausgeberin rekonstruieren dessen ‚wahren‘ Text.
- (3) Forscher oder Forscherin interpretieren ihn.
- (4) Leser oder Leserin werden durch diese Interpretation erleuchtet.

²⁵ Zu einer stärker dem Handwerklichen verpflichteten Betrachtung der sich bei Verwendung der digitalen Medien ergebenden Änderungen für die Rolle des Herausgebers siehe Claire Warwick: „Rumours of my death have been greatly exaggerated“: Scholarly Editing in the Digital Age. In: *New Media and the Humanities: Research and Applications*. Hg. von Domenico Fiormonte und Jonathan Usher. Oxford 2001. S. 49–56.

Wichtig für unsere Argumentation ist dabei, wie schon erwähnt, dass jeder dieser Schritte voraussetzt, dass der vorherige abgeschlossen wurde, bevor man ihn tut, und dass er nicht zurückgenommen werden kann. Falls Leser oder Leserin sich durch eine historische Interpretation im Dunkeln gefühlt haben, können sie sich brieflich beschweren: Selbst die brillianteste persönliche Beantwortung ihrer Fragen durch den Verfasser verbessert die Verständlichkeit des einmal geschriebenen Textes für Dritte jedoch nicht. Falls ein inhaltlich arbeitender Forscher auf der Basis seiner Kenntnisse des weiteren Entstehungskontextes gute Gründe findet, eine zentrale Lesung zu ändern, so hilft das ihm, die nächste Leserin sieht sich jedoch immer noch der unkorrigierten ursprünglichen Edition gegenüber.

Der wesentliche Unterschied bei einer Edition, die als „Prozess“ verstanden wird, ist, dass alle diese Schritte weitgehend unabhängig voneinander erfolgen,²⁶ also schon in dem Moment, in dem eine Quelle digital im globalen Informationsangebot „sichtbar“ wird, beginnen können. Im folgenden gehen wir also davon aus, dass alle Aktivitäten (spätestens nachdem Archivar oder Bibliothekarin ihre Tätigkeit beendet haben, aber durchaus auch schon, während sie sukzessive einen größeren Bestand bereitstellen) gleichzeitig ablaufen.

- (1) Archivar oder Bibliothekarin digitalisieren das Dokument.
- (1) Herausgeber oder Herausgeberin *beginnen* mit der Rekonstruktion des ‚wahren‘ Textes.
- (1) Forscher oder Forscherin interpretieren zunächst das unbearbeitete Dokument.
- (1) Leser oder Leserin erfreuen sich an der Ästhetik des Gesehenen.

Zumindest für bekennende Prä-Poststrukturalisten, wie den Verfasser, ist die Freude am nicht verstandenen Dokument gegenüber der Erleuchtung durch eine wohl erwogene historische Analyse geringer zu werten. Dem steht aber gegenüber, dass nach diesem Modell alle am Prozess Beteiligten sofort profitieren – nicht erst nachdem das Dokument, gefiltert durch das Lebenswerk eines Herausgebers und das weitere Lebenswerk eines interpretierenden Historikers, seinen Weg zum interessierten Leser gefunden

²⁶ Ein frühes Konzept für eine rechnergestützte Arbeitsumgebung, die eine derartige Vorgehensweise unterstützt: Better Access to Manuscripts and Browsing of Images. Aims and Results of an European Research Project in the Field of Digital Libraries: BAMBI Lib-3114. Hg. von Andrea Bozzi. Bologna 1997.

hat. Und diese Dynamik ist allen dabei das ursprüngliche Dokument mit interpretierendem Wissen anreichernden Tätigkeiten zu eigen. Editorische Arbeit schlägt sich in unserem Diagramm durch die durch (2) gekennzeichnete ‚Phase‘ in der Bearbeitung.

- (2) Archivar oder Bibliothekarin digitalisierten das Dokument.
- (2) Herausgeber oder Herausgeberin bieten erste, vorläufige Kommentare.
- (2) Forscher oder Forscherin sind in der Lage, auf Grund der editorischen Vorarbeiten die erste Interpretation zu verbessern.
- (1) Leser oder Leserin erhalten erste Hilfen zum Verständnis des rätselhaften Schönen.

Natürlich setzt sich diese ‚unmittelbare‘ Wirkung des intellektuellen Bemühens auf einer Ebene für die Voraussetzungen der Arbeit auf der nächsten weiter. So kann zwar die folgende unglückliche Situation entstehen:

- (3) Archivar oder Bibliothekarin digitalisierten das Dokument.
- (3) Herausgeber oder Herausgeberin boten erste, vorläufige Kommentare.
- (3) Forscher oder Forscherin bieten eine Darstellung für die Laien.
- (3) Leser oder Leserin verstehen sie nicht.

Der Vorteil eines dynamischen Konzepts auf der Basis elektronischer und jederzeit modifizierbarer Darstellungen beruhender Bereitstellung von Interpretation erlaubt es den interpretierenden Forschern jedoch, nicht nur brieflich auf die Verwirrung ihrer Leser und Leserinnen einzugehen, sondern die elektronisch publizierte Darstellung selbst zu ändern. Was das folgende Bild ergibt:

- (4) Archivar oder Bibliothekarin digitalisierten das Dokument.
- (4) Herausgeber oder Herausgeberin boten erste, vorläufige Kommentare.
- (4) Forscher oder Forscherin verbessern die Darstellung für die Laien.
- (4) Leser oder Leserin werden erleuchtet.

4. Ein ernst gemeinter Hinweis

Ich sagte einleitend, es schiene mir unter anderem deshalb sinnvoll, Ranke an den Beginn einer Darstellung dieser Art zu setzen, weil seine nur schein-

bar bescheidene Aussage aus einer Phase der Entwicklung der historischen Wissenschaften stamme, in der die historischen Fächer im Bewusstsein ihrer unbestrittenen Bedeutung ruhten. Dass es eine Phase gab, in der dies der Fall war, scheint mir angesichts des defensiven, ja manchmal geradezu wehleidigen Selbstbildes, das die historischen Disziplinen in heutigen forschungspolitischen Debatten geben, erinnerungsbedürftig.

Für die Gründe dieses Rückganges der Wertschätzung der Disziplinen gibt es viele Gründe; und nicht alle davon sind negativ, wenn wir bedenken, wie viel Prestige der Historiker aus seiner Rolle als Quasipriester einer pseudoreligiösen Verehrung für die eigene nationale Vergangenheit bezog (auch dann, wenn er selbst dieser nationalen Vergangenheit wesentlich rationaler gegenüberstand).

Doch einer der Gründe für die Wertschätzung der historischen Disziplinen lag in der Vergangenheit sicher darin, dass sie als die Schule für die kompetente Interpretation schwieriger Dokumente galten.

Die historischen Disziplinen werden mit den Möglichkeiten der Bereitstellung digitaler Informationen konfrontiert. Einer Fülle an widersprüchlichen, ungenauen, unübersichtlichen und lückenhaften Informationen, wie sie das heutige Internet prägen und wie sie die aus den penibel strukturierten, sauberen, widerspruchsfreien und vollständigen Informationswelten der Naturwissenschaften stammende Informatik oft eher ratlos lässt.

Es wäre schön, wenn die historischen Fächer, die ja seit langem Spezialisten für das Ungenaue, Unübersichtliche, Lückenhafte und Widersprüchliche sind, diese Fülle neuer Informationsmedien als Herausforderung begreifen, die energisch anzunehmen ihnen durchaus eine wichtige Rolle bei der Lösung der derzeitigen Probleme im Umgang mit den neuen Informationswelten zuweisen könnte. Denn: Wenn wir oben beschrieben haben, dass klassische Rollen in der neuen Situation überdacht werden müssen, so sind die historischen Disziplinen mit dieser Notwendigkeit keineswegs allein. Die Herausforderung zu übersehen, käme dagegen einem weiteren Rückzug in die Nische gleich.

5. Ein nicht ganz so ernst gemeinter Hinweis

Abschließend sei ein Hinweis darauf gestattet, dass die systematische Bereitstellung von Quellen aus Archiven und Bibliotheken einen willkomme-

nen Nebeneffekt haben könnte, nämlich die starke Vermehrung des Interesses an den Inhalten der bewahrenden Einrichtungen. Es ist unbestritten, dass es eine der ersten und vornehmsten Aufgaben des Archivars und Bibliothekars ist, das ihm Übergebene zu schützen. Vielleicht ist die dadurch bedingte geringe Zugänglichkeit der Inhalte von Archiven und Bibliotheken aber zu einem nicht geringen Teil mitverantwortlich daran, dass ihr Prestige in den letzten Jahren sank.

Als Beleg dafür, dass es ein erhebliches öffentliches Interesse auch an Dokumenten gibt, die der herrschenden Meinung nach nur dem Wohlgebildeten unmittelbar zugänglich und dem großen Rest der Bevölkerung nur durch ausgefeilte didaktische Präsentationen vermittelbar sind, lassen Sie mich einige Zitate aus einer Diskussion im Internet anführen, auf die ich stieß, als ich im Rahmen des oben zitierten Projektes zur Evaluierung der Retrodigitalisierungsaktivitäten abzuschätzen versuchte, *wer* eigentlich die Manuskripte, die unter <http://www.ceec.uni-koeln.de> zugänglich sind, in so erstaunlich großer Zahl betrachtete.

Festhalten möchte ich, dass die im Folgenden erscheinende Gruppe von Interessenten und Interessentinnen *keinen* plausiblen Grund für die digitale Bereitstellung von archivischer- und bibliothekarischer Überlieferung bildet. Wenn deren Interessen aber mit abgedeckt werden, so dienen wir mit dieser digitalen Bereitstellung vielleicht einer viel weiteren Öffentlichkeit, als wir selbst annahmen.

Somit zum leicht augenzwinkernden Ausklang, als Zitate, die für sich selbst sprechen sollen:

Mailing 1: 15.06.2001, 12:09

„Guten Tag!

Unziale? Karolingische Minuskel? Textura? Initiale? Illuminationen?

Wer schon immer mal mittelalterliche Handschriften aus der Nähe betrachten wollte, wird sich über die CODICES ELECTRONICI ECCLESIAE COLONIENSIS freuen: Die Universität Köln hat damit begonnen, die Dom- und Diozesanbibliothek zu digitalisieren. Neben allen für die verschiedenen Fachdisziplinen notwendigen Informationen können die kompletten Handschriften in vier Aufloe-

sungsstufen betrachtet werden. Wer's lesen will: Lateinkenntnisse sind nicht hinderlich.

<www.ceec.uni-koeln.de>

Es gruesst

Manfred“

Mailing 2: 19.06.2001, 08:43

„Hallo Manfred!

Auch wenn ich kein Latain (sic!) kann, die Seiten haben mich sehr beeindruckt. Facsimilie und Nachdrucke gibt es bei uns in Karlsruhe nämlich immer nur im Präsenzbestand, und hier kann man sich die Anregungen direkt auf den PC laden.

Wirklich toll. Dankeschön!

Es grüßt ganz herzlich

Ulrike“

Mailing 3: 19.06.2001 , 12:00

„Hallo Ulrike!

Und die haben quasi eben erst angefangen.

Ich ärgere mich manchmal sanft, daß ich Latein-Unterricht und Germanistik-Studium links liegen ließ, um als dummer Motorrad-Proll in den Tag hinein zu leben. Jetzt erfasse ich den Sinn der meisten Codices nur bruchstückhaft.

Ich hab mal Druckvorlagenhersteller und damit auch Schriftsatz gelernt und freue mich immer, wenn ich Beispiele aus der Schriftgeschichte finde. Das hier jedoch kommt derart geballt, daß ich schon eine Nacht damit verbracht habe, Alphabete der verschiedenen Autoren zusammenzustellen.

Es ist schon ein Unterschied, von Alkuin in der Übersetzung zu lesen oder direkt zu sehen, daß er sich nicht so wahnsinnig für Karls des Großen von oben verordnete Schreibreform interessierte. Da sehe ich Parallelen zu heute.

Im Herbst will ich damit beginnen, pseudoreligioese Schriften aus Clanthon im Codex-Stil festzuhalten (quasi ein Fanzine in Auflage 1). Tinten und Kiele habe ich schon, am Beschreibstoff mangelt's noch ...

Es gruesst

Manfred“

Datenstandards in der Erschließung historischer Dokumente

Patrick Sahle

1. Vorbemerkung

Der diesem Text zugrunde liegende mündlich-visuelle Tagungsvortrag hatte vor allem einführenden, orientierenden, fast schon Workshop-Charakter. Im Wechsel der situativen, kommunikativen und medialen Umstände wird er hier auf die wesentlichen theoretischen Grundlinien reduziert. Die einzelnen Standards werden nicht eingehend besprochen und auch das abschließende Praxisbeispiel wird nicht wiedergegeben: Anhand des Projektes *Zentrales Verzeichnis Digitalisierter Drucke (zvdd)* waren im mündlichen Vortrag die Auswahl von Standards, ihre Lokalisierung im vorgestellten allgemeineren Modell und die praktischen Schwierigkeiten ihrer Anwendung besprochen worden. Bei dem hier vorliegenden Text handelt es sich inhaltlich eher um ein Abstract, umfänglich um ein ausformuliertes, erweitertes Abstract.

2. Zur Bedeutung von Standards

Jeder Fall ist ein Sonderfall. Alle Gegenstände der historischen Überlieferung erfordern ihre je eigenen Weisen der Beschreibung und Erschließung. Jede Materialgattung, jede Überlieferungssituation und alle jeweils zu erwartenden Auswertungsinteressen könnten die anzuwendenden technischen Standards in einer Weise bestimmen, die für jedes Projekt zu einer individuellen Lösung führen müsste. Dieser strikt material- und situationsorientierte Ansatz muss aber derzeit aus einer ganzen Reihe von Gründen

immer mehr relativiert werden. Mindestens drei Prinzipien des gegenwärtigen Medienwandels haben als allgemeine Paradigmen weit reichende Auswirkungen auf alle Teilbereiche der konzeptionellen und praktischen Konfiguration der Erschließung. Die Rede ist hier

- vom inkrementellen Prinzip,
- vom Prinzip der Vernetzung und
- vom Prinzip der Transmedialisierung.

Was bedeuten diese Prinzipien für die ‚Erschließung‘, unter der in einem weiten Begriffsverständnis alle Arbeiten zusammengefasst werden können, die historisch überlieferte Materialien durch Wiedergabe, kritische Beschreibung und Informationsanreicherung für die weitere wissenschaftliche Verwendung vor- und aufbereiten?

Früher war ein Faksimile ein Faksimile, ein Katalog ein Katalog, ein Regestenband ein Regestenband, eine Edition eine Edition und eine wissenschaftliche Monographie eine wissenschaftliche Monographie. Abgeschlossene und für sich stehende Informationseinheiten. Abgeschlossene Publikationsformen, die zwar immer auch aufeinander verwiesen, sonst aber in keinem weiteren direkten Verhältnis zueinander standen. Heute – realistisch betrachtet aber in der allgemeinen Praxis wohl erst morgen – ist alles potentiell vernetzt. Das, was uns früher als expliziter Verweis erschien, wenn zum Beispiel eine Edition sich auf die in den Quellenstudien, Katalogen und Findbüchern nachgewiesenen Materialien bezog, hat heute den Status nur eines impliziten Verweises. Schließlich ist mit der realen Verlinkung zwischen digitalen Ressourcen bzw. mit der Einbindung von digitalen Datenfragmenten eine neue Form der expliziten Verweisung und Bezugnahme zur leitenden Form des Möglichen geworden.

Die einzelnen Formen der Erschließung bauen in einer veränderten Weise explizit aufeinander auf. Dies lässt die Grenzen zwischen den Publikationseinheiten, die medial bedingt waren aber konzeptionell gefasst wurden, verschwimmen. Die traditionellen Formen greifen über sich hinaus: Der elektronische Katalog schließt auf der einen Seite digitale Faksimiles ein und bildet auf der anderen Seite potentiell auch eine Plattform für die Forschungsinstrumente zur Tiefenerschließung der Überlieferung. Die wissenschaftliche Auswertung, die geschriebene Geschichte, inkludiert ihre Grundlagen und macht sich so potentiell in einer neuen Weise transparent.

Auf der einen Seite verschwimmen die Grenzen. Auf der anderen Seite müssen die einzelnen Arbeiten an der Überlieferung und die dabei entstehenden Informationen doch wieder konzeptionell getrennt werden. Sie müssen in einer neuen Weise ‚modularisiert‘ werden. Sie müssen als Bausteine in einem umfassenden Informationsraum gedacht werden, der jetzt als in sich geschlossener und zusammenhängender Informationsraum eben dadurch besser sichtbar wird, dass er nicht mehr fragmentiert, sondern potentiell hoch integriert ist.

Die ‚Trennungen‘ in der digitalen Erschließung sind teilweise andere Trennungen als in der traditionellen Erschließung. Zunächst ist für die digitalen Daten oft die Rede von der Trennung von Inhalt und Form. Dabei handelt es sich um ein weit verbreitetes Grundkonzept der gegenwärtigen technischen Lösungsansätze. Gemeint ist, dass Informationen auf einer Ebene abstrakter Datenmodelle und Daten gespeichert werden sollen. Dabei sollen aber die Inhalte von Informationen gespeichert und verwaltet werden und nicht ihre äußere Erscheinungsform in jenen Präsentationsmedien, mit denen der menschliche Benutzer schließlich in Kontakt kommt. Durch die Speicherung verallgemeinerter, codierter Informationen wird es außerdem möglich, dass gleiche Daten in verschiedenen medialen Konfigurationen mit unterschiedlichen Ergebnissen algorithmisch ausgegeben werden. Das Konzept beinhaltet so die Trennung von transmedialen Daten und (beliebigen) medialen Ausgabeformen. Da die medialen Ausgabeformen erstens algorithmisch erzeugt werden und zweitens höchst variabel sind, geht es vordringlich nur um die Daten, die ‚Inhalte‘ selbst. Dies ist die oben angesprochene ‚Transmedialisierung‘.

Die transmedialisierten Inhalte können leichter explizit aufeinander aufbauen – sie können sich gegenseitig übernehmen und integrieren. Alle Arbeiten an der Überlieferung erscheinen heute als Bausteine einer Verarbeitungskette, die bei der unmittelbaren materiellen Überlieferung ansetzt und diese immer tiefer erschließt, kritisch sichtigend, kontextualisierend, transkribierend, deutend, analysierend immer weiter verarbeitet. Damit führen aber auch alle Arbeiten zu Bausteinen der Erschließung – zu Bausteinen, deren Wert darin liegt, dass sie möglichst einfach in anderen Zusammenhängen benutzbar sein sollten.

Wenn man nun betrachtet, wie die einzelnen Bereiche der Erschließung differenziert werden können, dann sehen wir, dass zunächst durchaus von den traditionellen Konzepten ausgegangen werden kann, um zu einer neu-

en Modularisierung zu kommen. Offensichtlich gibt es verschiedene Formen der Grunderschließung, die von der Objektbeschreibung über die Katalogisierung bis zur Regestierung (gewissermaßen einer ersten inhaltlichen Durchdringung) reichen kann. Neu ist hier, dass noch ganz am Anfang die digitale Faksimilierung durch die Umkehrung der traditionellen Kostenrelationen¹ zu einer Selbstverständlichkeit wird, die zu einer anderen Transparenz der weiteren Erschließungsschritte führt und die relativen Positionen der einzelnen Module neu bestimmt.²

Auf die Grunderschließung setzt die Textgewinnung auf. Hier wird der Text zunehmend als Skala verschiedenster, im Idealfall algorithmisch und dynamisch zu generierender Textformen erkennbar, die zwischen den Polen quellennaher (benutzerferner) und quellenferner (benutzernaher) Darstellungsweisen eine Vielzahl von Positionen einnehmen können. Auswertende Formen der Arbeit mit der Überlieferung können in gleicher Weise auf einer Skala fortschreitender Verarbeitung und zunehmender Distanz zu den Quellen verortet werden. Auf dieser Verarbeitungskette können wir durchaus den gleichen Akteuren einen Platz zuweisen, die wir auch in der traditionellen Form spezielle Arbeiten durchführen gesehen haben: Archivare und Bibliothekare kümmern sich vor allem um Grunderschließungsleistungen. Wissenschaftliche Editoren erstellen verschiedene Formen von Textrepräsentationen. Und Fachwissenschaftler der unterschiedlichsten Richtungen werten das überlieferte Material schließlich für spezielle Fragestellungen aus. Allein die Übergänge werden jetzt vielleicht noch fließender als sie es bislang schon waren, da die integrative Kraft der transmedialen Daten die Grenzen zwischen den Publikationsformen dieser Arbeitsschritte verwischen.

Die Modularisierung der Arbeitsschritte geschieht auf der Achse der zunehmenden Verarbeitung der Überlieferung. Sie bezieht sich auch auf die Rolle der verschiedenen Akteure im Prozess der umfassenden Erschließung. Sie ergibt sich aber auch aus einem anderen Aspekt der gegenwärtigen Informationstechnologien: Vernetzung, Integration und Transmediali-

¹ Darauf war bereits Manfred Thaller in seinem Eröffnungsreferat eingegangen.

² In dem Moment, in dem das digitale Faksimile selbstverständlich als erstes verfügbar gemacht wird, verändert sich der ursprüngliche ‚Stellvertretercharakter‘ von Katalogisaten, Quellenbeschreibungen, Regesten und Textausgaben. Sie werden nun stärker als Informationsdestillate, als Hilfsmittel zum Auffinden von Quellen und als vorverarbeitende Formen für spezielle Auswertungsstrategien erkennbar.

sierung können nur dann funktionieren, wenn klar ist, welches die Objekte sind, die wir inhaltlich repräsentieren wollen. Die gegenwärtigen Informationstechnologien sind darauf angewiesen, mit klar definierten Informationseinheiten zu operieren. Auf unterschiedlich granularen Ebenen muss klar sein, wovon jeweils die Rede ist, was eigentlich der – zunächst konkrete, dann aber auch fortschreitend abstrakt gefasste – Gegenstand ist, von dem die Daten handeln. Auch deshalb ist eine klare Modularisierung unseres Informationsraumes unerlässlich. Wir müssen absehen können, über welche Objekte Informationen zu erwarten sind, wie wir diese (repräsentierten) Objekte ansprechen können und wie wir Informationen zu ihnen verstehen und verarbeiten können. Die Form definierter Objekte und der Zuschnitt der Informationen zu ihnen müssen berechenbar sein, um in anderen Kontexten benutzt werden zu können. Und sie müssen bekannt sein, damit Schnittstellen zwischen technischen Systemen funktionieren können, über die Daten weitergegeben werden können.

Die Definition von Objekten, Teilbereichen der Verarbeitung, Festlegung von Sichten auf Objekte und Schnittstellen ist auch deshalb so wichtig, weil die Organisation des Informationsraumes tatsächlich noch ein wenig komplexer ist, als dies bislang angedeutet worden ist. Es muss nämlich weiterhin unterschieden werden zwischen der Eingabeform von Informationen, der Form der Datenhaltung und der Ausgabeform. Erschließungsinformationen können auf verschiedenen Wegen gewonnen werden. Durch die Konversion bereits vorhandener Ressourcen ebenso wie über die Eingabe neuer Informationen. Die Schicht der Datenhaltung ist davon weitestgehend unabhängig. Zwischen beiden besteht ein Verhältnis der Übersetzung und Abstraktion. Auf der anderen Seite besteht ein Verhältnis der Übersetzung und medialen Konkretisierung zwischen der Verwaltung allgemeiner Daten und ihren Darstellungsweisen in den verschiedenen medialen Konfigurationen. Erschließungsdaten können hier ebenso zu einem gedruckten Werk ausformatiert werden, wie mit ihnen unter Umständen eine interaktive Netzpublikation angetrieben werden kann. Was die Modularisierung des Informationsraumes betrifft, so können auf der Seite der Publikation verschiedene inhaltliche Module wieder zusammenfließen.

Eine weitere Differenzierung der Teilbereiche kann noch im Bereich der Datenhaltung vorgenommen werden. Hier ist – für manche technische Ansätze – noch zwischen Grunddaten, Strukturinformationen und Metadaten zu unterscheiden.

Der Informationsraum der Erschließung historischer Dokumente lässt sich insgesamt wie folgt skizzieren:

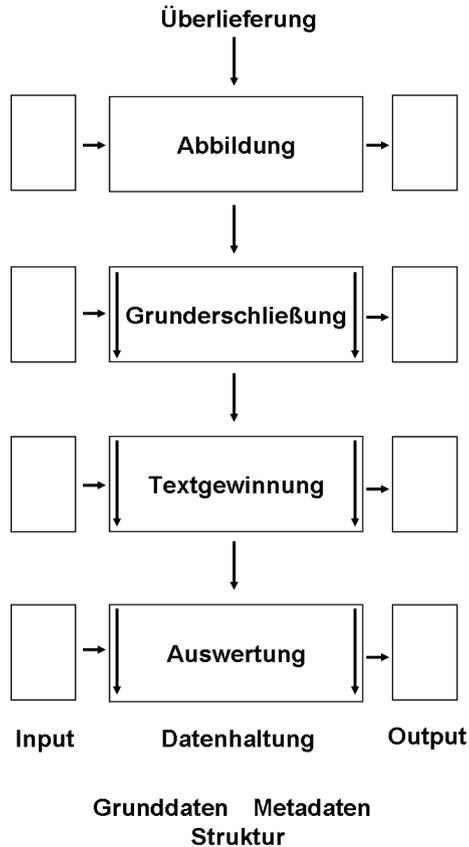


Abbildung: Erschließung der Überlieferung als Informationsraum.

Das Modell des Informationsraumes soll dabei helfen, die Bedeutung von konzeptioneller Modularisierung, klarer Objektdefinitionen, von Standards und von Schnittstellen in der Erschließung der historischen Überlieferung einsichtig zu machen. Es dient aber auch als Folie, um Standards zu lokalisieren bzw. allgemeiner, um einzelne Erschließungsprojekte und deren Verwendung von Standards zu beschreiben.

Die Bedeutung technischer Standards liegt in der Ermöglichung eines modularen Erschließungskonzeptes. Die Grundfunktion von Standards ist es, Informationen durch Gleichmäßigkeit der Beschreibung auch nach außen besser verstehbar und berechenbar zu machen. Diese Berechenbarkeit ist die Grundlage für die Erfüllung weiterer Anforderungen an digitale Erschließungsinformationen: für die Nachhaltigkeit und langfristige Nutzbarkeit der gewonnenen Daten, für die algorithmische weitere Verarbeitung in anderen Nutzungssituationen und für die Anschlussfähigkeit und Integrationsfähigkeit in weitere Kontexte. Denn die einzelnen ‚Bausteine‘ der Erschließung werden dadurch über ihre unmittelbare Erstellungssituation hinaus wertvoll, dass sie auch später noch in anderen Zusammenhängen nachgenutzt, weiter verarbeitet oder in andere Systeme integriert werden können. Dazu gehören dann zum Beispiel vertiefende Erschließungsarbeiten. Dazu gehören unter Umständen aber auch Suchverbünde oder Portale, die Teilinformationen herausziehen, um die erschlossenen Objekte von einer übergeordneten Warte aus nachzuweisen, besser auffindbar und besser nutzbar zu machen.

3. Standards im Informationsraum

Im Bereich der Standards begegnen einige scheinbare Paradoxien. So gibt es zwar einerseits bereits eine große Fülle von Standards, andererseits aber passt selten ein Standard ganz genau auf die konkret vorliegenden Probleme. Man soll die vorhandenen Standards benutzen und nicht immer wieder neue Standards erfinden, muss dann aber immer einen Mittelweg finden zwischen der glatten Anwendung bestehender Standards und den eigenen Bedürfnissen. Diese führen in der Regel zu Anpassungen, die zwar in so genannten ‚Anwendungsprofilen‘ dokumentiert werden können, letztlich aber den Status der verwendeten Datenbeschreibung ‚als Standard‘ gefährden. Dieses Problem resultiert daraus, dass Standards und die Situationen, in denen sie verwendet werden, durch eine ganze Reihe von Aspekten bestimmt werden, die noch über die Modularität des erschließenden Informationsraumes hinaus das Feld der Datenbeschreibungen ausdifferenzieren.

Für die Lokalisierung und Beschreibung von Standards, aber auch für die Auswahl zur Verwendung in bestimmten Projektkontexten, werden eine Zahl von Parametern genutzt, die im Folgenden aufgelistet werden.

3.1 Grad der Allgemeinheit

Standards können sich auf unterschiedlichen Abstraktionsebenen befinden. Das ‚relationale Datenmodell‘ oder das ‚hierarchische Datenmodell‘ sind allgemeine Weisen der strukturierten Abbildung von Informationen. Diese können sich in konkreteren Standards niederschlagen: SQL (Standard Query Language) ist ein Abfragestandard für Daten, die dem relationalen Modell folgen. XML (eXtensible Markup Language) ist eine Realisation des allgemeinen Modells ‚Textauszeichnung‘. Dieses ist selbst sowohl an sequentiellen wie auch hierarchischen Grundstrukturen orientiert. XML ist ein Beispiel für einen Meta-Standard, der aus einer allgemeinen Warte zwar konkret ist (indem er ein bestimmtes Datenmodell realisiert und genaue Struktur- und Syntaxvorgaben macht), aus einer speziellen Warte aber allgemein. Für die konkrete Anwendung muss auf der Grundlage des Meta-Standards erst eine Anwendungssprache formuliert werden.

3.2 Determinismus

Standards sind wie Sprachen. Ihr Funktionieren liegt weniger in einer abstrakten Vorgabe, wie etwas gemeint ist, als vielmehr im gemeinsamen Gebrauch der Sprecher. In der Regel sind Standards nicht selbsterklärend, sondern müssen dokumentiert werden, um ihren Sinn zu klären. Weil diese Dokumentation aber natürlichsprachlich erfolgt, bleibt notwendigerweise immer Raum für abweichende Interpretationen bei den Benutzern. Dieser Raum kann unterschiedlich groß sein. Ein Standard kann zu sehr gleichmäßigen Anwendungen führen wie auch zu sehr ungleichmäßigen. Dies hängt nicht unbedingt von der Präzision und Vollständigkeit der Dokumentation ab, sondern auch von den anderen hier beschriebenen Parametern.

3.3 Beschreibung oder Austausch

Standards können die Beschreibung von Informationen festlegen oder einen technischen Rahmen für den Austausch von Daten bilden. Solche Schnittstellen-Standards können wiederum unterschiedlich allgemein sein.

Allgemeine Übertragungsprotokolle wie HTTP (Hypertext Transfer Protocol) oder Schnittstellen wie CGI (Common Gateway Interface) bilden nur einen allgemeinen Rahmen für die Kommunikation der Daten, ohne etwas über die Inhalte auszusagen. Speziellere Schnittstellen-Protokolle wie Z39.50 bzw. ZING (Z39.50 International Next Generation) oder OAI-PMH (Open Archives Initiative, Protocol for Metadata Harvesting) betreffen dagegen bestimmte Datenbereiche, hier zum Beispiel die Metadatenschicht.

3.4 Materialgattung

Offensichtlich brauchen verschiedene Materialien ihre eigenen Standards zur Beschreibung. Die TEI-Guidelines (Text Encoding Initiative) betreffen Volltexte, der MASTER-Standard die Codierung von Handschriftenkatalogen (Manuscript Access through Standards for Electronic Records).³ EAD (Encoded Archival Description) bezieht sich auf Findmittel in Archiven. SVG (Scalable Vector Graphics) dient der Wiedergabe von ‚Bildern‘ (Skizzen, Diagramme), die sich über Vektoren beschreiben lassen.

3.5 Verarbeitungsgrad

In meinem Modell des Informationsraumes spielte die fortschreitende Verarbeitung eine entscheidende Rolle. Manche Standards betreffen eher Grunddaten, manche spätere Verarbeitungsstufen. MASTER ist ein Standard für die Katalogisierung von Text-Objekten, TEI ein Standard für die Wiedergabe der Texte *in* den Objekten.

3.6 Input – Datenhaltung – Output

Manche Standards überspannen den Dreischritt Eingabe – Verwaltung – Ausgabe. Oft ist aber die Verwendung unterschiedlicher Lösungen sinnvoll. Dabei besteht für die Verwendung von allgemeinen Standards auf der Ebene des Inputs noch die geringste Notwendigkeit: hier können durchaus lokale technische Lösungen verwendet werden, weil der Austausch und die weitere Verwendung der gewonnenen Daten ja erst auf der Ebene der Datenhaltung ansetzt. Allein in sehr hoch modularisierten Projekten mit zum Beispiel weit verteilter Datenerfassung ist die Verwendung von Standards

³ MASTER ist allerdings in der letzten Fassung der TEI-Guidelines (P5) aufgegangen.

bereits bei der Dateneingabe wichtig. Ein Standard für die Erfassung von Daten ist XForms (XML-Formulare). Die meisten Standards betreffen die Ebene der Datenhaltung, da es hier um eine möglichst sachadäquate, differenzierte Beschreibung und Verwaltung von Informationen geht. Datenhaltungs-Standards sind oft zugleich noch relativ allgemein, so dass sie für eine unmittelbare Darstellung in Output-Formen gar nicht geeignet sind. Dies betrifft zum Beispiel XML im Allgemeinen oder konkrete XML-Auszeichnungssprachen wie die TEI. Auf der anderen Seite können solche allgemeinen Daten dann algorithmisch in Ausgabeformate überführt werden, die so präzise sind, dass sie von verschiedenen Medien unmittelbar dargestellt werden können. Hier wäre zum Beispiel auf TEX, PS (PostScript) oder PDF (Portable Document Format) zu verweisen, mit denen Druckmedien generiert werden können. Mit PDF, dann aber vor allem auch (X)HTML (Hypertext Markup Language), wären hier auch die wichtigsten Standards für den Output von Texten in digitalen Medien genannt. Ähnliche Unterscheidungen wären parallel zum Beispiel für den Bereich der Bilddaten zu machen: Hier ist TIFF ein Eingabeformat, LZW-komprimiertes TIFF ein Speicherformat und JPEG, GIF oder PNG Ausgabeformate.

3.7 Informationsdimensionen

Grunddaten, Strukturdaten, Metadaten, Semantik. Standards können sich auf unterschiedliche Ebenen der digitalen Repräsentation, auf unterschiedliche Sichtweisen auf Objekte beziehen. ASCII (American Standard Code for Information Interchange) und UNICODE betreffen die Zeichencodierung von Textdaten, Auszeichnungssprachen wie die TEI regeln die Beschreibung der Struktur von Volltexten. Die TEI umfasst dann aber auch schon den Bereich der Metadaten, der für andere Objekte von gesonderten Standards abgedeckt wird: Sollen zum Beispiel nur die Metadaten bibliographischer Einheiten erfasst werden, so stehen dafür bibliothekarische Formate wie MAB (Maschinelles Austauschformat für Bibliotheken) oder MARC (Machine Readable Catalogue) zur Verfügung. Ein wichtiger Standard, der zwar für digitale Ressourcen im Allgemeinen ‚zuständig‘ ist, sich hier aber ebenfalls ausdrücklich auf die Metadatenschicht beschränkt, ist DC (Dublin Core). Metadaten- und Strukturdaten digitaler Objekte deckt METS (Metadata Encoding and Transmission Standard) ab, auf die Strukturdaten gedruckter Daten beschränkt sich E-Bind (electronic Binding)

DTD). Gerade in letzter Zeit hat sich jenseits der deskriptiven Metadaten noch eine Schicht der eher deutenden semantischen Daten etabliert. Auch bisher haben inhaltlich orientierte Beschreibungsstandards wie MIDAS (Marburger Informations-, Dokumentations- und Administrationssystem) für Objekte der Kunst oder ‚Iconclass‘ für die Klassifikation von Bildinhalten im Grunde eine semantische Deutung ihrer Gegenstände vorgenommen. Mit den Standards im Vor- und Umfeld des kommenden ‚Semantic Web‘ sind hier aber allgemeinere Formate entstanden, die semantische Inhalte und Bezüge von digital beschriebenen Objekten verwaltbar machen. Zu nennen wären hier zum Beispiel RDF (Resource Description Framework), XTM (XML Topic Maps) oder OWL (Web Ontology Language).

3.8 Konzeptionalisierung

Selbst wenn für einen bestimmten Anwendungsfall alle bereits genannten Parameter klar sein sollten, so führt dies nicht dazu, dass am Ende *ein* Standard als Mittel der Wahl zur Verfügung stehen würde. Oft würde man feststellen, dass für den speziellen Fall noch kein Standard entwickelt worden ist, so dass man auf eine Beschreibungsweise ausweichen müsste, die in dem einen oder anderen Parameter von den eigentlichen Anforderungen abweicht. Oft hat man aber auch die Auswahl zwischen mehreren Standards. Diese unterscheiden sich dann zum Beispiel in der Art und Weise, wie sie die Beschreibung der Objekte konzeptionalisiert haben. Ein Beispiel hierfür wären MAB und MARC: Beide betreffen die Metadaten zu bibliographischen Einheiten. Sie beschreiben sie aber durchaus mit leicht unterschiedlichen Konzepten.

3.9 Technische Umsetzung

Selbst wenn sonst alle anderen Parameter übereinstimmen sollten, so kann ein Unterschied zwischen verschiedenen Standards außerdem immer noch in der technischen Umsetzung auch gleicher Konzepte liegen. Dabei können die technischen Unterschiede Auswirkungen auf die Gewinnung, die Speicherung und Verwaltung, den Austausch und die Darstellung der Informationen haben und dadurch Kriterien für die Benutzung eher des einen oder des anderen Formates liefern. Ein Beispiel aus diesem Bereich wären die Standards zur Codierung von Bildern nach dem Bildpunkte-

Prinzip („bitmaps“). JPEG (Joint Photographic Experts Group), PNG (Portable Network Graphics), GIF (Graphics Interchange Format) oder TIFF (Tagged Image File Format) dienen alle der Codierung von Bildern nach dem gleichen Prinzip. Sie unterscheiden sich allein nach der Art der Datenkomprimierung, den Farbräumen und den Möglichkeiten zusätzlicher Dokumentation der Bilder. Bereits dies können aber Aspekte sein, die im jeweiligen Anwendungsfall den Einsatz eines bestimmten Formats nahe legen.

4. Probleme und Fehlerquellen bei der Verwendung von Standards

Die Auswahl und Anwendung geeigneter Standards ist nicht trivial. Standards können mehrere Felder abdecken, sich überschneiden, andere Standards beinhalten oder die Verwendung ergänzender Standards nahe legen. Weiterhin ist zu überlegen, ob und wie weit Standards für die lokalen Erfordernisse angepasst werden sollten. Bei der Auswahl und Anwendung von Standards treten oft ähnliche Probleme auf, werden häufig ähnliche Fehler gemacht. Auf einige sei hier abschließend hingewiesen.

4.1 Falsches Datenmodell

Zu den seltener werdenden Fehlern gehört, dass für Informationen, die nun einmal eine bestimmte Grundstruktur aufweisen, das falsche Grunddatenmodell gewählt wird. In der Vergangenheit ist es häufig vorgekommen, dass zum Beispiel hierarchisch organisierte Dokumentinformationen oder Textdaten mit einem relationalen Modell abgebildet wurden.

4.2 Nicht materialadäquater Standard

Manchmal werden Standards für allgemeiner gehalten als sie es tatsächlich sind. TEI ist ein Standard, der vornehmlich aus literaturwissenschaftlicher und linguistischer Sicht heraus die Textgattungen dieser beiden Teilfächer beschreiben sollte. Wird TEI für andere Textgattungen (zum Beispiel Urkunden) und andere Analyseabsichten (zum Beispiel historische Forschung) eingesetzt, dann muss mit einer Schiefelage bei der Anwendung ge-

rechnet werden. Ein Standard ist auch ein WahrnehmungsfILTER, mit dem die betrachteten Dinge in einem ganz bestimmten Licht erscheinen.

4.3 Anpassung und Dokumentation

Zuweilen muss beim Fehlen eines geeigneten spezialisierten Standards auf einen anderen Standard zurückgegriffen werden, der eigentlich nicht ganz genau passt. Erfolgt dann keine Anpassung an die speziellen Gegebenheiten oder aber keine Dokumentation der veränderten Benutzung, dann gerät die Hauptfunktion der Standards, nämlich berechenbare Beschreibungen zu erzeugen, in Gefahr. Man wird dann in der Weiterverwendung entweder Daten bekommen, die nicht der Sache entsprechen oder Daten, die zwar der Sache entsprechen, deren Funktionsweise aber nicht mehr dem Ausgangsstandard entspricht und deshalb nicht berechenbar ist.

4.4 Vermischung von Datenhaltung und Output

Einer der häufigsten Fehler, die auch heute noch immer wieder begegnen, ist die Verwendung von Output-Formaten für die Datenhaltung. Immer wieder gibt es Projekte, die allgemeine Daten nicht nur in PDF oder HTML darstellen, sondern diese Formate auch als zentrale Organisations- und Datenhaltungsstandards benutzen. Das ist dann legitim, wenn man entweder außer der unmittelbaren Publikation alle weiteren Verwendungsmöglichkeiten für die Daten von vornherein ausschließen kann oder wenn man der Auffassung ist, dass die zu erschließenden Objekte wirklich keine weiteren Informationen enthalten als jene, die sich auch mit den Publikationsformaten wiedergeben lassen. Wenn es zum Beispiel um Texte geht, dann ist ein Standard wie PDF als Datenhaltungsstandard nur unter der einschränken- den Prämisse vernünftig einsetzbar, dass man unter ‚Text‘ keine weiteren Informationen versteht als jene, die sich auf der typographischen Oberfläche abbilden lassen. Zu bedenken ist aber auch, dass Standards für die Medialisierung oft recht stark an den gegenwärtigen Hard- und Software- umgebungen und an den gegenwärtigen Rezeptionsgewohnheiten orientiert sind. Dies steht tendenziell im Widerspruch zu einer langfristigen Nutzbarkeit, die auf möglichst allgemeinen Formen der Datencodierung und Datenstrukturierung beruhen sollte.

4.5 Mapping

Ein letztes Problem betrifft die Verwendung von Inhalten aus verschiedenen Standards. Daten müssen in der Regel gemappt werden, um sie in anderen Kontexten weiter verwenden zu können. Dabei handelt es sich um einen ganz klassischen Übersetzungsprozess, mit den bekannten Problemen der Übersetzung: Eine rein mechanische Übersetzung, die einfach von der Definition des jeweiligen Standards ausgeht, wird zwangsläufig Informationsverluste und Fehlinterpretationen nach sich ziehen. Es ist oft unumgänglich, die jeweiligen Daten und ihre Entstehungsbedingungen wenigstens auf der Projektebene etwas genauer zu verstehen, um sie sinnvoll weiter benutzen zu können. Standards bieten hier die Grundlage, um Daten überhaupt effizient ‚mappen‘ zu können. Sie bieten aber oft keine Garantie für das Funktionieren einer unmittelbaren, unkontrollierten und undifferenzierten Verarbeitung

5. Schluss

Der vorliegende Beitrag hat versucht, das Feld der Verwendung von Standards bei der Erschließung historischer Dokumente zu umreißen. In einem allgemeinen Modell der fortschreitenden Erschließung als einem umfassenden Informationsraum können einzelne Module isoliert werden, denen verschiedene technische Standards entsprechen. Standards können aber auch unter weiteren Aspekten beschrieben und verortet werden. Diese können innerhalb konkreter Forschungsprojekte bei der Auswahl und Anwendung geeigneter Standards ebenso helfen wie bei der Suche nach möglicherweise nicht optimalen Verwendungsweisen.

Fachspezifische Indexierung von historischen Dokumenten I

Quellen zwischen Zeichenketten und Information – Beispiel Urkunden

Georg Vogeler

1. Einleitung

Die Geschichtswissenschaft digitalisiert sich: Es werden Portale aufgebaut, fachspezifische Metasuchen installiert, Onlinerezensionen und -publikationen verbreiten sich und werden immer häufiger zitiert (wenn auch vorwiegend innerhalb der ‚Gemeinde‘). Die ‚Cultural Heritage‘-Institutionen wagen sich immer weiter in die digitale Welt, an ihrer Spitze die Bibliotheken, aber dicht gefolgt von Museen und Archiven. Es ist also an der Zeit, sich zu fragen, wie man die von diesen Institutionen bereitgestellten historischen Informationen auch wiederfindet.

Neben den allgemeinen Suchmaschinen wie Google, Yahoo oder Ask gibt es zu diesem Zweck auch fachspezifische Portale und Suchmaschinen. *Chronicon* ist ein solches Angebot, das die Bayerische Staatsbibliothek aufgebaut hat.¹ Es vermittelt den Zugang zu 32 bibliographischen Onlinehilfsmitteln, die mit einer einzigen Suchabfrage gemeinsam abgefragt werden können. Auch das Berliner Portal *Clio Online* bietet eine solche Metasuche an: Hier gibt es neben der Suche in 28 bibliographischen Datenbanken auch ein Metaportal zur Quellenrecherche.² Wenn man sich die dahinter stehenden Datenbanken genauer ansieht, stellt man fest, dass es sich um genau

¹ Adresse: <http://www.chronicon.de/> (letzte Einsichtnahme am 21.04.2006).

² Adresse: <http://www.clio-online.de/> (letzte Einsichtnahme am 21.04.2006).

drei handelt: das Archivportal des Landes Nordrhein-Westfalen,³ die VD17--Datenbank der Herzog-August-Bibliothek in Wolfenbüttel⁴ und die Nachlassdatenbank des Bundesarchives.⁵

Ist es etwa so, dass die Grundlage historischer Arbeit, die Quellen, noch gar nicht online suchbar ist? Sicherlich sind im Netz noch nicht so viele Quellenbestände recherchierbar wie mit gedruckten Hilfsmitteln. Die Situation ist jedoch beileibe nicht schlecht. So haben andere Archivverwaltungen auch ähnliche Angebote wie das Land Nordrhein-Westfalen eingerichtet: das Portal des Landesarchivs Baden-Württemberg,⁶ das Hessische Archiv-Dokumentations- und -Informationssystem *HADIS*⁷ oder *Ariadne*, das *Archive Information and Administration Network* der Archive in Mecklenburg-Vorpommern.⁸ In diesen Portalen kann man sich einerseits zu den einzelnen Archiven durchklicken, man kann aber auch in allen Bestandübersichten und Findbüchern, die schon digital vorliegen, suchen – ohne erst das einzelne Archiv anzusteuern. Wir haben es also je mit einer Art ‚Quellensuchmaschine‘ zu tun.

Neben den Archiven sind auch andere Quellencorpora online zu finden. Ich möchte hier keine umfangreiche Übersicht über solche Corpora und ihre unterschiedlichen Präsentationsformen bieten, sondern kurz zwei Angebote nennen, die sich mittelalterlichen Urkunden verschrieben und sie auf besondere Art und Weise durchsuchbar gemacht haben: Da sind die hessischen *Landgrafen-Regesten online*⁹ und der *Codice diplomatico della Lombardia medievale*.¹⁰ Das Erstere fällt durch die Möglichkeit auf, in den Volltexten nach lemmatisierten Wörtern zu suchen, das heißt Varianten von Wörtern in Flexionsformen durch Angabe des Grundwortes, das Letztere durch eine umfangreiche Expertensuche, die einen tief in die sachlichen und sprachlichen Strukturen des Textes hineinführt.

³ Adresse: <http://www.archive.nrw.de/> (letzte Einsichtnahme am 21.04.2006).

⁴ Adresse: <http://www.vd17.de/> (letzte Einsichtnahme am 21.04.2006).

⁵ Adresse: <http://www.nachlassdatenbank.de/> (letzte Einsichtnahme am 21.04.2006).

⁶ Adresse: <http://www.landearchiv-bw.de/> (letzte Einsichtnahme am 21.04.2006).

⁷ Adresse: <http://www.hadis.hessen.de/> (letzte Einsichtnahme am 8.04.2006).

⁸ Adresse: <http://ariadne.uni-greifswald.de/index.html> (letzte Einsichtnahme am 21.04.2006).

⁹ Adresse: http://online-media.uni-marburg.de/ma_geschichte/lgr/ (letzte Einsichtnahme am 18.04.2006).

¹⁰ Adresse: <http://cdlm.unipv.it/> (letzte Einsichtnahme am 21.04.2006).

Ein Konzept für eine fachspezifische Suchmaschine, die dem Historiker dabei hilft, Quellen zu seinen Fragestellungen zu suchen, ja vielleicht sogar nicht nur Quellenreferenzen, sondern auch schon Quellenaussagen, muss sich mit den Fähigkeiten solcher Suchen vergleichen lassen.

Worin liegt denn der Vorteil der beiden Suchen? Sie ermöglichen es dem Historiker, Informationen zu finden, ohne die genaue Zeichenrepräsentation dieser Information zu kennen: Die Suche nach Kaufgeschäften in den hessischen Regesten kann mit einer Suchen nach „kaufen“ gestartet werden und muss sich nicht darum kümmern, ob das Phänomen „kaufen“ mit einem Partizip „gekauft“ oder in Flexionen wie „kauft“ ausgedrückt ist.

Die online verfügbaren Editionen des *Codice diplomatico della Lombardia medievale* reichen bis zum Jahr 1150. Ein Historiker, der sich für Anwesenheiten, Aktivitäten und Rezeption der Ottonen in Norditalien interessiert, kann sich zu einer Personensuche vorarbeiten. Dort kann er die 45 Einträge zu „Otto imperator“ ebenso aussuchen wie die 3 zu „Otto I., imperatore“, „Otto II., imperatore“ und zu „Otto III., imperatore“. Eine Volltextsuche nach „Otto“ (676 Treffer) liefert gänzlich andere Ergebnisse: Ottonis (555 Treffer) wird bei einer einfachen Stichwortsuche übergangen, bei einer rechtstrunkierten Suche dafür auch „Ottobelli“ (38 Treffer) mit ausgewiesen. Die ottonischen Kaiser sind also nur als Personen und nur schwierig als Teil des Volltextes identifizierbar.¹¹

Die mittelalterlichen Urkunden werden auch weiterhin der Testfall sein, an dem die vorhandenen Suchmaschinentechnologien exemplifiziert werden. Wie weit die für dieses Material sinnvollen Techniken auch für andere Quellengattungen nützlich sind, ist weiteren Studien vorbehalten.

2. Suchmaschinentechnologie

Eine moderne Suchmaschine hat – sehr vergrößert – etwa folgenden Aufbau: Ein Crawler durchsucht das Internet, das heißt er verfolgt alle auf Webseiten vorhandenen Links. Die dabei gefundenen Webseiten übergibt er an Software, die das Dokument in einzelne Wörter zerlegt und daraus einen Index aufbaut. Wenn der Index abgefragt wird, dann wird die gefun-

¹¹ Es sei noch darauf hingewiesen, dass auch die Personenindexeinträge voneinander abweichen: Otto imperator: 39, Otto imperator augustus: 2, Otto imperatore: 2, Otto, imperator: 2, Otto I, II, III, je ein Eintrag.

dene Menge an Treffern nach Relevanz geordnet. Die Techniken zum Suchen der Dokumente, die Organisation eines schnellen Index und die Methoden zur Bewertung der Relevanz eines Dokumentes für die Suchanfrage sind die drei Bereiche, in denen Suchmaschinenhersteller ihre ganz spezifischen Kompetenzen und Geschäftsgeheimnisse haben. Eine fachspezifische Suchmaschine muss sich eben den selben drei Bereichen widmen. Hier soll es nun vorrangig um den Index gehen, auch wenn einige Überlegungen zur Bewertung mit einfließen.

Aus technischer Sicht ist der Index die Form, wie die Daten so abgelegt werden, dass die Maschine schnell auf sie zugreifen kann. Die einschlägigen Stichwörter lauten ‚B-Baum‘, ‚Hash‘, ‚R-Baum‘ usw. und meinen verschiedene Arten, die einzelnen Wörter so auf der Festplatte abzulegen, dass ein gegebenes Suchwort schnell mit ihnen verglichen werden kann. Die Zugriffsgeschwindigkeit ist dabei die Kernfrage. Welcher Art die Daten sind, interessiert nicht, denn sie sind einfach ein beliebiger Code. Anders formuliert: Die Techniker interessieren sich vorrangig dafür, wie der Computer schnell eine beliebige Zeichenkette in einer Liste findet. Ob diese Liste Personennamen, Archivsignaturen, Berufscodes einer Volkszählung oder Genomsequenzen enthält, interessiert nicht.

Historiker haben da eine andere Perspektive. Sie denken an das Register im Buch, das ihnen hilft, eine bestimmte Stelle im Text zu finden, ein Kapitel, einen Abschnitt, der sich mit Friedrich II. beschäftigt oder die Urkunden, die für den Deutschen Orden ausgestellt worden sind. Sie beginnen die Recherche in einem Quellencorpus im Register, wenn sie den Kontext des Wortes „gubernare“ ermitteln wollen oder wenn sie etwas über die Mühlen im 13. Jahrhundert erfahren möchten.

Quellen sind damit für Historiker nicht nur reine Zeichenketten, sondern Informationsträger. Sie brauchen Indices, die an diesen Informationen orientiert sind. Eine Volltextsuche ist dabei nur ein Hilfsmittel bei der Informationsrecherche. Eine fachspezifische Suchmaschine muss also einen Index aufbauen, der möglichst informationsnah gestaltet ist.

Dabei kann es nicht darum gehen, dass eine Heerschar an Wissenschaftlern alle online verfügbaren Quellenbestände systematisch erschließt. Es muss im Gegenteil darum gehen, Quellenbestände, die ohne Erschließung ins Netz gestellt werden, mit Hilfe von kalkulierbaren Prozeduren automatisch aufzubereiten, so dass sie ihre individuellen Eigenheiten behalten können, im Index der Suchmaschine aber eine Repräsentanz finden, die

mehr über ihren Informationsgehalt verrät als nur eine Liste aller Zeichenketten, die zwischen Worttrennungselementen, das heißt Leerräumen, Satzzeichen usw. stehen.

3. Möglichkeiten zur graphischen und sprachlichen Abstraktion

Eine erste Abstraktionsschicht zwischen Quellentext und Information stellt die Orthographie dar. Wo wir in den jüngsten Diskussionen um ‚alte‘, ‚neue‘ und ‚reformierte‘ Rechtschreibung gelernt haben, dass es sogar in einer orthographisch sehr regelbewussten Zeit Varianten gibt, die gleiche Wörter unterschiedlich graphisch repräsentieren, so sind an Quellen arbeitende Historikerinnen und Historiker sich der um so größeren Vielfalt bis ins 17. Jahrhundert wohl bewusst. Ein erstes Ziel einer automatisierten Abstraktion von den Zeichenketten der digitalen Quellenrepräsentanz ist es also, möglichst viele dieser Varianzen auszuschalten bzw. zu tolerieren.

Die Techniken dazu sind nichts Neues. Sie sind auch schon in komplexe fachspezifische Softwarelösungen eingeflossen.¹² Sie seien hier aber erneut zusammengestellt und auf ihre Nutzbarkeit für eine fachspezifische Suchmaschine geprüft.

Eine grundlegende Technik, graphische oder sprachliche Varianzen in eine Suche einzufügen, sind die ‚regulären Ausdrücke‘. Diese in vielen Softwareumgebungen implementierten Mustervergleiche bieten eine Notation an, eine Zeichenkette aus Alternativen bestehen zu lassen. Ein solcher Ausdruck „\<O[dt][.]{0,1}o.*?\>“ sucht nach Wörtern, die mit einem „O“ anfangen, auf das ein „d“ oder ein „t“ folgt, dem noch maximal ein zweiter Buchstabe folgen kann, ein „o“ folgen muss, bevor das Wort beliebig enden kann. Damit lassen sich zum Beispiel Schreibungsvarianten des Namens „Otto“ abbilden: „Oto“, „Odo“, „Oddo“, „Oddonis“, „Othonem“ entsprechen alle diesem Muster.

¹² Vgl. Manfred Thallers *kleio* (Adresse: <http://www.hki.uni-koeln.de/kleio/>, letzte Einsichtnahme am 20.04.2006) und darauf aufbauende Anwendungen. Einem dem hier vorgestellten Ansatz vergleichbar ist: Andrea Ernst-Gerlach und Norbert Fuhr: Generating Search Term Variants for Text Collections with Historic Spellings. In: 28th European Conference on Information Retrieval Research (ECIR 2006). Adresse: http://www.is.informatik.uni-duisburg.de/bib/pdf/ir/Ernst_Fuhr:06.pdf (letzte Einsichtnahme am 21.04.2006), die jedoch nicht bei der Indexierung, sondern beim Ranking ansetzen.

Einem ähnlichen Prinzip folgen ‚Name Matching‘-Algorithmen, von denen es eine größere Menge gibt. Der bekannteste ist das schon 1918 entwickelte ‚Soundex‘-Verfahren, das den amerikanischen Einwanderungsbehörden half, der Sprachen- und Namensvielfalt der europäischen Einwanderer Herr zu werden. Der klassische Soundex-Algorithmus erhält den ersten Buchstaben eines Namens und fügt bis zu 3 Ziffern an, die für die Konsonanten des Namens stehen und sie in Gruppen ordnen. Aus „Ottomem“ wird dann „O355“; aus „Odo“ „O300“ und aus „Othonis“ „O352“. Das Verfahren ist relativ stark auf die englische Sprache ausgerichtet, weshalb verschiedene Modifikationen vorgeschlagen worden sind.¹³

Es ist sichtbar, dass auch eine solche Umwandlung immer noch ungenaue Ergebnisse liefert. Ungenauigkeit ist in der Informatik auch aus der Arbeit mit automatischer Texterkennung bekannt: Leichte Abweichungen vom orthographisch eingeführten Zeichenbestand entstehen dabei leicht. Um auch in solchen ‚unsauberen‘ Texten suchen zu können, sind Algorithmen entwickelt worden, die sich dem Problem des ‚nearest neighbour‘ widmen. Dabei spielt der Grad der Veränderung eine entscheidene Rolle: Am geläufigsten ist wohl die 1965 vorgestellte ‚Levenshtein-Distanz‘, die die Anzahl der mindestens notwendigen Operationen beziffert, die es braucht, um aus einer Zeichenkette eine zweite werden zu lassen. Damit kann ein Wert festgelegt werden, bis zu dem zwei Zeichenketten, die nur leicht voneinander abweichen (zum Beispiel in genau einem Zeichen) als gleich gewertet werden.¹⁴

Aus historischer Perspektive erscheint es unsinnig, beliebige Änderungen gleich zu behandeln: Es ist offensichtlich, dass bestimmte Abweichungen zweier Zeichenketten für einen gegebenen sprachlichen Kontext weniger auffällig sind als andere: Bis ins 17. Jahrhundert kann man davon ausgehen, dass „u“ und „v“ nur positional und nicht lautlich differenziert verwendet werden. Eine Abweichung von „vnnß“ zu „uns“ als pure Zeichenkette ist mit einer Levenshtein-Distanz von 3 zu bewerten, obwohl „u“ und „v“, „ß“ und „s“ sowie „nn“ und „n“ für den geübten Leser als iden-

¹³ Einen guten Überblick bis zum Stand 1995 gibt: A.J. Lait und B. Randell: An Assessment of Name Matching Algorithms [1995]. Adresse: <http://homepages.cs.ncl.ac.uk/brian.randell/home.-informal/Genealogy/NameMathing.pdf> (letzte Einsichtnahme: 31.03.2006), die auch gleich eine weitere Variante („Phonex“) beisteuern.

¹⁴ Der Algorithmus wurde veröffentlicht: Vladimir I. Levenshtein: Binary codes capable of correcting deletions, insertions, and reversals. In: Doklady Akademii Nauk SSSR 163 (1965). S. 845–848 (Russisch). Englische Übersetzung in: Soviet Physics Doklady 10 (1966). S. 707–710.

tisch zu bewerten sind. Für das Niederdeutsche hat Jan Strunk einmal daraus die Konsequenz gezogen und einen Algorithmus vorgeschlagen, der diese in der Abstandsberechnung unterschiedlich gewichtet und insbesondere durch Berücksichtigung graphischer Varianzen signifikant bessere Suchergebnisse erzielt.¹⁵

Im oben angeführten hessischen Beispiel kommt jedoch keine dieser Techniken zum Einsatz. Statt dessen verwendet die Anwendung ein Verfahren, das aus der Computerlinguistik als ‚Stemming‘ bekannt ist. Dahinter verbirgt sich eine morphologische Lemmatisierung, die Flexionsformen eines Wortes auf ihr Grundwort zurückführt. Auch für das Lateinische existiert Software zur morphologischen Analyse, wie zum Beispiel ‚Lem-Lat‘¹⁶ oder das ‚Perseus Word Study Tool‘.¹⁷ Diese Ansätze basieren gewöhnlich auf Lexika und Regelwerken zur Flexion, Präfix- und Suffixbildung. Damit verwendet diese Technik nicht nur inhärente Strukturen der Texte, die algorithmisch aufbereitet werden, sondern externes Wissen, das in Thesauri abgelegt ist. Dennoch sind die Verfahren des Stemming keine reinen Wortlisten, die ganze Flexionsparadigmen ihrem Grundwort zuordnen, sondern Wörterbücher der Grundwörter mit formalisierten Angaben zur Morphologie, so dass der Computer jede mögliche Flexionsform des Wortes bilden kann.

Es gibt aber auch Anwendungen, die ganze Wörter in Referenzlisten zusammenführen. Insbesondere prosopographische Arbeit hat Namenslisten erstellt, die leider nur selten öffentlich zugänglich sind: So sind die Register der jüngeren Bände des *Repertorium Germanicum* mit solchen Namenslisten entstanden. Auch die Namensdatenbanken, die im Zusammenhang mit dem *Altdeutschen Namenbuch* und dem Projekt *Nomen et gens* entstanden sind, sind nicht öffentlich verfügbar.¹⁸ Es ist offensichtlich noch nicht wahr-

¹⁵ Jan Strunk: Information retrieval for languages that lack a fixed orthography. Adresse: <http://www.linguistics.ruhr-uni-bochum.de/~strunk/LSreport.pdf> (letzte Einsichtnahme am 21.04.2006).

¹⁶ Adresse: <http://webilc.ilc.cnr.it/~ruffolo/> (letzte Einsichtnahme am 20.04.2006).

¹⁷ Adresse: <http://www.perseus.tufts.edu/cgi-bin/morphindex?lang=la> (letzte Einsichtnahme am 20.04.2006).

¹⁸ *Nomen et gens*: Adresse: <http://www.nomen-et-gens.de/portal.asp> (letzte Einsichtnahme am 20.04.2006). Vgl. auch Redmer Alma: A multi-level database. Study on the elite of 15th and 16th century Groningen. In: Prosopography and Computer. Contributions of Medievalists and Modernists on the Use of Computer in Historical Research. Leuven, Apeldoorn 1995. S. 185–194. – Didier F. Isel: Prosopographie des personnages mentionnés dans les textes pour l'époque de

genommen worden, dass diese Materialsammlungen nicht nur zur Vorbereitung onomastischer und prosopographischer Forschung dienen, sondern ein nutzbringendes Hilfsmittel darstellen können, Quellenaussagen aus den Zeichenketten ihrer digitalen Repräsentation zu ermitteln. Eine fachspezifische Suchmaschine muss also den Versuch unternehmen, derartige Datenbestände zu nutzen. Es ist jedoch um vieles leichter, ‚Name Matching‘-Algorithmen und approximative Verfahren beim Erstellen eines Index zu verwenden, da hier vorhandene oder einmal zu entwickelnde Algorithmen direkt in die Softwarearchitektur eingebaut werden, statt umfangreiche Zugriffe auf bislang noch nicht einmal öffentliche externe Datensammlungen zu organisieren.

Für eine historisch orientierte Suchmaschine wären demnach Algorithmen zu entwickeln. Das Untersuchungsobjekt ‚mittelalterliche und frühneuzeitliche Urkunden‘ würde zum Beispiel besonders zwei Verfahren erfordern, eines für die graphischen Varianten des Mittellatein und eine für die des Frühneuhochdeutschen.

Das *DEEDS*-Projekt, das umfangreiche mittellateinische Urkundentexte englischer Klöster digitalisiert hat, verwendet dafür ein existierendes Verfahren, den ‚Double Metaphone‘-Algorithmus, der einen echten phonetischen Ausgleich versucht.¹⁹ Erfahrungen damit liegen derzeit noch nicht vor. Auf Grund der Erfahrungen von Jan Strunk, der die Qualität der Suchergebnisse insbesondere durch Verfahren auf graphischer Ebene verbessert sieht, würde ich einen einfachen ‚Graphex‘-Algorithmus vorschlagen, der folgende Regeln berücksichtigen würde: „h“ wird nicht berücksichtigt. „cio“ und „tio“ werden gleichbehandelt, „e“ „ae“ „oe“ und e-caudata werden gleichbehandelt, ebenso „i“, „y“ und „ii“. Für die Epenthese (zum Beispiel „damnum“ = „dampnum“) und die Assimilation (zum Beispiel „auctor“ > „autor“) werden Ausgleichsklassen gebildet.

Auch für das Mittelhochdeutsche wie für das Frühneuhochdeutsche existieren noch keine Ausgleichsalgorithmen. Ein ‚Graphex‘ dazu müsste die Konsonantenreduplikationen berücksichtigen, „v“ und „u“ sowie „ß“

Pépin le Bref et de son frère Carloman (741–768) spécialement ceux exerçant une fonction ecclésiastique ou laïque, Quatrième édition (mars 2006): Tableau de concordance entre noms et lemmes (Adresse: <http://prosopographie-id.de/Tableau%20de%20concordance.htm>).

¹⁹ Vgl. zum Algorithmus: Lawrence Phillips: The Double Metaphone Search Algorithm. In: C/C++ User Journal June 2000. Adresse: <http://www.cuj.com/documents/s=8038/cuj0006philips/> (letzte Einsichtnahme am 31.03.2006).

und „s“ gleichbehandeln, „cz“ mit „z“ gleichsetzen und die Schreibweisen „ew“ und „aw“ mit den Diphthongen „eu“ und „au“ identifizieren.²⁰

Die Realität dialektal eingefärbter Urkundensprache des 14.–16. Jahrhunderts würde auch ‚Soundex‘-ähnliche Ausgleichsmechanismen nahelegen: p=b, mb=b, d=t, g=k=c=q sind geläufige lautliche Verwandtschaften. Die deutsche Sprachgeschichtsforschung kennt noch eine größere Menge zeitlich und geographisch begrenzter Abweichungen,²¹ kann aber ohne ein gerade erst im Entstehen befindliches Corpus²² noch keine genaueren Aussagen darüber treffen, wann wo welche Varianten auftreten. Ob also die Berücksichtigung dieser weitergehenden Varianten nutzbringend ist, bedarf genauer Evaluation.

Mit derartigen Ausgleichsmechanismen besteht die begründete Hoffnung, dass die Stichwörter aus dem aktuellen Wortschatz eines Historikers, in die er sein Forschungsproblem übersetzt und die er in eine mit diesem Mechanismen ausgestattete Suchmaschine eingibt, eine Sammlung auch orthographisch und lautlich varianter Quellentexte ermittelt, die Auskunft zu seiner Frage geben, dass er also zum Beispiel Quellendokumente findet, die eine historische Schreibweise des gesuchten Personennames enthalten.

4. Informationstypen

Für historische Arbeit ist darüber hinaus aber noch eine Eigenschaft von Quellentexten zentral, die in einer Suche, die nur auf Zeichenketten orientiert ist, nicht möglich ist: die zeitliche Einordnung. Zeitliche Einordnung beruht nämlich auf numerischen Angaben. Für eine zeitliche Eingrenzung einer Suche auf Quellenmaterial aus der Zeit Ottos III. wären Quellentexte zu filtern, die Datumsangaben aus der Zeit zwischen 996 und 1002 enthal-

²⁰ Eine Beispielimplementation für den mittellateinischen und den frühneuhochdeutschen Graphex und einen oberdeutschen Soundex in Visual Basic for Applications findet man unter Adresse: <http://www.cei.lmu.de/examples/graphex.bas.zip> (letzte Einsichtnahme am 01.05.2006)

²¹ Vgl. zum Beispiel Frédéric Hartweg und Klaus-Peter Wegera: Frühneuhochdeutsch. Eine Einführung in die deutsche Sprache des Spätmittelalters und der frühen Neuzeit (Germanistische Arbeitshefte 33). 2. überarb. Aufl. Tübingen 2005; Gerhard Philipp: Einführung ins Frühneuhochdeutsche. Sprachgeschichte, Grammatik, Texte. Heidelberg 1980.

²² Vgl. *Deutsch Diachron Digital*: Adresse: <http://www2.hu-berlin.de/ddd/> (letzte Einsichtnahme am 20.04.2006) mit Links auf einzelne Teilprojekte (Adresse: <http://www2.hu-berlin.de/ddd/links.php>).

ten. Ein Ausdruck „>995 und <1003“ führt in alphanumerischer Ordnung, das heißt also in einer Ordnung rein nach Reihenfolge der einzelnen Zeichen, zu keinem Ergebnis, denn es gibt kein Zeichen, das in alphanumerischer Reihung hinter „9“ und vor „1“ liegt. Ebenso sind Abweichungen, die über approximative Suchverfahren ausgeglichen werden, anders zu beurteilen, indem die Änderung einer Stelle am Beginn einer Zahl eine weit größere Abweichung darstellt als am Ende einer Zahl. Ein fachspezifischer Index muss also zwischen alphanumerischen und numerischen Anfragen unterscheiden.

Zunächst erscheint die Unterscheidung leicht, da für beide Informationstypen bestimmte Zeichen reserviert sind: Ziffern und Buchstaben. Ich möchte das Problem gemischter Formen, lateinischer Zahlen und sprachlicher Ausdrücke für Zahlenwerte beiseite lassen und das Problem ‚Zahl/Wort‘ als allgemeineres Problem von Informationstypen weiterverfolgen.

Viele der bislang angesprochenen Techniken erscheinen im historischen Alltag überflüssig: Um einen Bibliothekskatalog benutzen zu können, braucht der Forscher eine bibliographisch regelgerechte Referenz, die im 20. Jahrhundert zum Standard wissenschaftlichen Arbeitens geworden ist. Eine Fußnote oder eine wissenschaftliche Bibliographie liefern mir also ausreichend Informationen, eine Angabe gemäß der strikten Regeln des Kataloges zu finden, ja sie sogar für meine Suche nutzbar zu machen, indem klare Ein- und Ausgrenzungen möglich sind. Auch die Suche im Findbuch im Archiv braucht in vielen Fällen keinen Ausgleich historischer Schreibweisen, denn es ist häufig genug von einem Zeitgenossen angelegt worden, der moderner Orthographie folgt – von den Findbüchern des 19. Jahrhunderts sei hier einmal abgesehen. Beide Informationsquellen haben jedoch in der Welt der Bibliotheken und des Internets eines gemeinsam: Sie sind ‚Metadaten‘, das heißt Angaben, die das eigentliche Objekt – das Archival, das Buch – beschreiben. Damit sehen sie grundsätzlich anders aus als der eigentliche Text einer Quelle und sind ‚Informationstypen‘ wie Zahlenangaben oder geographische Angaben mit eigenen Eigenschaften. Weitere geläufige Informationstypen findet man in der Welt der gedruckten Bücher in den Registern: Ortsindex, Personenindex oder Sachindex sind alles Texte, die eine bestimmte Information repräsentieren, die bei einer Suche unterschiedlich ausgewertet werden kann. Die Funktionen dieser Informationstypen sind in der elektronischen Welt über die Anwendungen sichtbar gemacht, in die sie eingebettet sind. So kann zum Beispiel der

Karlsruher Virtuelle Katalog Informationstypen in verschiedenen Datenbeständen zusammenfassen, indem er auf die standardisierten Schnittstellen der internationalen Bibliotheksanwendungen zugreift.²³ Eine fachspezifische Suchmaschine kann das nicht, da die Anwendungen zur Präsentation der historischen Dokumente keine solchen Schnittstellen kennen.

5. Strukturen

Dennoch kann sie die Informationstypen unterscheiden. Die technische Entwicklung im WWW entfernt sich von HTML, der ‚HyperText Markup Language‘, die dem Internet in 1990er Jahren sein äußeres Gesicht gegeben hat. Auch wenn die Diskussion um das ‚Semantische Netz‘ oder ‚Web 2.0‘ nicht mehr die Feuilletons beherrscht,²⁴ XML dominiert die technische Diskussion, wenn es um Publikationen im Internet geht.

Die Phänomene HTML und XML werden in der Informatik unter dem Stichwort ‚Semistrukturierte Daten‘ diskutiert. Darunter werden Phänomene zusammengefasst, die sich mit relationalen und objektorientierten Datenbanksystemen nicht oder nur ungünstig abbilden lassen, weil unübersichtlich komplexe Datenstrukturen entstehen und fehlende oder wiederholte Informationstypen umfangreiche Leerstellen entstehen lassen. Die Textauszeichnungssprachen HTML und XML sind darüber hinaus so erfolgreich, weil sie nicht nur unregelmäßig strukturierte Daten abbilden können, sondern auch weil sie ‚selbstbeschreibend‘ sind, das heißt, weil die strukturtragenden Daten Teil der Texte selbst sind.²⁵

Ein fachspezifische Suchmaschine wird also auf diesen Eigenschaften der Dokumente im Web aufbauen. Welche Konsequenzen das haben kann,

²³ Über den KVK: Adresse: http://www.ubka.uni-karlsruhe.de/hylib/virtueller_katalog.html (letzte Einsichtnahme 20.04.2006).

²⁴ Der Einführungsvortrag zur Tagung *.hist2003* widmete sich noch dieser Vision (vgl. Christoph Albrecht: Geschichte und Neue Medien. In: Geschichte und Neue Medien in Forschung, Archiven, Bibliotheken und Museen, Tagungsband *.hist2003*. Hg. von Daniel Burgkhardt und anderen (Historisches Forum 7). Berlin 2005. Adresse: http://edoc.hu-berlin.de/e_hist-for/7_I/PHP/Ueberblicke_7-2005-I.php#001001, letzte Einsichtnahme am 20.04.2006). Auf der Folgetagung *.hist2006* war das Stichwort jedoch aus den Titeln der Vorträge und Sektionen verschwunden (vgl. Adresse: http://www.clio-online.de/hist2006/texts/hist2006_programm.pdf, letzte Einsichtnahme am 20.04.2006).

²⁵ Vgl. zu den ‚Semistrukturierten Daten‘: Serge Abiteboul, Peter Buneman und Dan Suciu: Data on the Web. From Relations to Semistructured Data and XML. San Francisco 2000.

möchte ich an den mittelalterlichen und frühneuzeitlichen Urkunden deutlich machen und bei einem Vorhaben beginnen, das zunächst wie das Gegenteil einer Suchmaschine klingt, die automatisch historischen Texten wahrscheinliche sachliche Repräsentanzen zuordnet, ein Vorhaben, das jedoch Konzepte verdeutlicht, die in den Funktionsumfang einer solchen Suchmaschine nutzbringend eingebaut werden können.

Die *Charters Encoding Initiative* (CEI)²⁶ ist eine Arbeitsgruppe, die sich im Frühjahr 2004 das Ziel gesetzt hat, einen Vorschlag für einen Textauszeichnungsstandard für die Quellengattung ‚Urkunden des Mittelalters und der frühen Neuzeit‘ zu machen. Ein solches Bestreben beruht auf den Beobachtungen, dass erstens die Historiker, Archivare und Bibliothekare diese Quellengattung in größerem Maß digitalisiert und über das Netz zugänglich gemacht haben und dass zweitens die Urkunden strukturierter Erschließung besonders zugänglich sind.²⁷

Die Arbeitsgruppe geht auf einem induktiven und auf einem deduktiven Weg zu einem gemeinsamen Textauszeichnungsstandard. Induktiv sind folgende zwei Maßnahmen: Die Arbeitsgruppe ermittelt erstens Ähnlichkeiten vorhandener Schemata und analysiert zweitens die Arbeitsprozesse bei der Arbeit mit Urkunden. Dabei ist deutlich geworden, dass die an unterschiedlichen Stellen unabhängig von einander entwickelten Datenstrukturen auffällig ähnlich sind. Diese Ähnlichkeit beruht nicht nur darauf, dass natürlich immer das gleiche Objekt beschrieben wird, sondern sogar in den Namen der Strukturelemente finden sich Ähnlichkeiten: Für Abkürzungen verwenden zum Beispiel die den Regeln der TEI angelehnten Editionen der *École nationale des chartes*²⁸ ebenso <abbr> wie der *Codice diplomatico della Lombardia medievale*.²⁹

Die Eigenschaft von XML-Dokumenten, ‚selbstbeschreibend‘ zu sein, geht also teilweise so weit, dass unabhängig von einander entwickelte Datenstrukturen direkt übertragbar werden. Ein erstes Ziel der Arbeitsgruppe

²⁶ Adresse: <http://www.cei.lmu.de> (letzte Einsichtnahme am 20.04.2006).

²⁷ Vgl. Georg Vogeler und Patrick Sahle (Hg.): Virtual Library Historische Hilfswissenschaften, Sektion Diplomatik. Adresse: <http://www.vl-ghw.lmu.de/diplomatik.html> (letzte Einsichtnahme am 20.04.2006).

²⁸ Adresse: <http://elec.enc.sorbonne.fr> (letzte Einsichtnahme am 21.04.2006). Die XML-Daten dieses Projektes sind über die Webseite nicht einsehbar.

²⁹ Adresse: <http://cdlm.unipv.it> (letzte Einsichtnahme am 21.04.2006). XML-Daten können in diesem Projekt über die Suchfunktion (‚Mostra la marcatura XML‘) auch über die Webseite eingesehen werden.

ist es deshalb, einen gemeinsamen Wortschatz für die Auszeichnung von digitalen Urkundenrepräsentationen bereitzustellen.³⁰

Der deduktive Weg geht von einer internationalen Terminologie der Diplomatie aus.³¹ Das *Vocabulaire Internationale de Diplomatie* ist nämlich nicht nur ein Wörterbuch, das die in 12 europäischen Sprachen und im Lateinischen üblichen Fachausdrücke zur Beschreibung von Urkunden und ihrer Entstehung zusammenstellt, sondern bildet durch seinen systematischen Aufbau auch eine Struktur diplomatischer Informationen ab.³²

Eine fachspezifische Suchmaschine wird sich also an diesen Strukturen orientieren. Sie wird die existierenden Strukturen aus den Dokumenten auslesen und sie mit den vorhandenen Strukturen der CEI vergleichen. Obwohl zurzeit noch Handarbeit notwendig ist, erscheint es möglich, mit Hilfe des *Vocabulaire Internationale de Diplomatie* auch XML-Strukturen der Ausgangsdokumente automatisch auf einen Standard nach den Vorgaben der CEI abzubilden.

Dabei entstehen im Index der Suchmaschine natürlich sehr ungleiche Strukturen. Welche Bedeutung kann man dann den Strukturen noch zu-messen? Zunächst einmal wird aus den existierenden Beispielen deutlich, dass eine fachspezifische Suchmaschine in Datenbeständen sucht, nicht in Webdokumenten. Das heißt die Suchmaschine wird versuchen, das einzel-ne Dokument aus der Struktur der Datei zu filtern und dafür die Treffergenauigkeit auszuwerten.

Eine ganz neue Perspektive ergibt sich, wenn man sich an Diskussionen im Bereich ‚Information Retrieval‘ heranwagt. Neben Überlegungen, auch die Semantik eines Textes mit in die Suche einzubeziehen, die neben den oben angeführten Verfahren unter dem Stichwort ‚Textmining‘ komplexe Konzepte entwerfen, die Bezüge und Bedeutungen von Wörtern formal ab-zubilden, konzentriert sich ein Strang der Diskussion auf die Strukturen.³³

³⁰ Vgl. den Vorschlag unter Adresse: <http://www.cei.lmu.de/taglibrary.html> (letzte Einsichtnahme am 21.04.2006).

³¹ *Vocabulaire international de la diplomatie*. Hg. von Maria Milagros Cárcel Ortí (Collecció Oberta). 2. verbesserte Auflage. València 1997.

³² Vgl. dazu auch Michele Ansani: *Diplomatica (e diplomatismi) nell'arena digitale*. In: *Archivio storico italiano* 158 (2000). S. 349–379, auch in: *Scrineum* 1 (1999). S. 1–11 (Adresse: <http://dabc.unipv.it/scrineum/ansani.htm>, letzte Einsichtnahme am 05.03.2000).

³³ Vgl. Norbert Fuhr: *Information Retrieval Methods for Literary Texts*. In: *Jahrbuch für Computerphilologie* 5 (2003). S. 147–160 (online unter Adresse: <http://computerphilologie.uni-muenchen.de/jg03/fuhr.pdf>, letzte Einsichtnahme am 21.04.2006).

Welche Bedeutung die Strukturen für die Bewertung eines Treffers haben können, kann gut am Urkundenbeispiel erläutert werden. Volkssprachliche mittelalterliche Urkunden sind nämlich nicht nur Quellen für Historiker, sondern auch hochgeschätzte Quellen für die Sprachgeschichte, die in ihnen regional und zeitlich klar zugeordnete Textzeugen findet. Das Suchinteresse des Sprachhistorikers ist so in vielen Fällen ein gänzlich anderes: Zwar sucht er auch nach Urkunden, aber er sucht nach Urkundentexten, nicht nach Urkunden als Zeugnissen von Sachverhalten. Ja sogar die oben angeführten Verfahren, variante Darstellungen von Wörtern mit ihren modernen Formen in Übereinstimmung zu bringen, sind für den Sprachhistoriker nicht notwendigerweise neue Zugänge zu bislang verschlossenen Informationen, sondern sie können ebenso auch Unterscheidungen überdecken, die zentral sind für die Anfrage.

Für ein konkretes Projekt einer fachspezifischen Suchmaschine können diese unterschiedlichen Suchinteressen durch unterschiedliche Anfrageseiten berücksichtigt werden.³⁴ Es kann aber auch eine andere Erkenntnis daraus verallgemeinert werden: Moderne Suchausdrücke für historische Sachverhalte sind in den Originaltexten seltener und ungenauer zu erwarten als in den Metadaten, die sprachlich normalisiert sind. Eine historische Anfrage an ein digitales Urkundencorpus wird deshalb Treffer in den Kopfregersten höher bewerten als Treffer im Editionstext. Die aus XML-Dokumenten ermittelbaren Datenstrukturen liefern also nicht nur Hinweise auf die Datentypen und die darauf anzuwendenden Werkzeuge graphematischer, phonetischer und lexikalischer Konkordanz, sondern auch Hinweise auf die Relevanz eines Treffers für die Anfrage.

6. Fazit

Zunächst sei festgestellt, dass die angesprochenen Instrumente keineswegs eindeutige, positivistische Sachaussagen aus den Quellentexten werden ermitteln können. Die Frage- und Deutungskompetenz der Historiker stellt eine fachspezifische Suchmaschine nicht in Frage. Sie liefert ihm ein Hilfs-

³⁴ Vgl. Georg Vogeler: Europäisches Urkundenerbe. Zu Potentialen und Perspektiven eines internationalen Fachinformationssystems digitaler Urkundenpublikationen. In: Elektronische Fachinformationssysteme in der Geschichte, Jahrestagung der AGE, 25.–26.11.2006. Hg. von Franz Götz. München 2006 (im Druck).

mittel an die Hand, das Heterogenitäten der Quellencorpora verringert – und damit die Menge der relevanten Dokumente erhöht (recall) – und die Relevanz der Treffer für eine historische Anfrage besser sichtbar macht – und damit den Anteil an irrelevanten Treffern, die der Suchende verarbeiten muss, verringert (precision).

Eine derartige Suchmaschine muss aber mit zwei Problemen fertig werden:

1. Die oben beschriebenen Werkzeuge entfalten ihre Wirkung erst, wenn sie auf passende Texte angewendet werden: Ein für frühneuhochdeutsche Texte getrimmter ‚Graphex‘ wird auf altfranzösische Texte angewendet kaum sinnvolle Ergebnisse liefern. Sowohl die Implementierung des im Eingang dieses Textes erwähnten aber nicht weiterverfolgten Crawlers, wie die Analyse der gefundenen Texte, müssen darauf geprüft werden, ob es Instrumente gibt, die Sprache der Texte plausibel zu bestimmen. Im hier verfolgten Testscenario mittelalterlicher und frühneuzeitlicher Urkundencorpora haben wir es noch nicht mit den Materialmengen zu tun, die dazu zwingen, diesen Schritt zu automatisieren. Vorläufig wird die Suchmaschine mit den Corpora von Hand gefüttert, das heißt vor der Indexierung des Corpus kann der Bearbeiter entscheiden, welche Art von Transformationen angewendet werden soll.
2. Die Vielzahl der Techniken ist hier weitgehend theoretisch beschrieben. Es gibt einzelne Evaluationen der vorgestellten Techniken,³⁵ aber viele sind auch noch nicht evaluiert, ja noch nicht einmal implementiert. Das sind insbesondere die Techniken, die sich mit mittelalterlichen und frühneuhochdeutschen Sprachständen europäischer Volkssprachen und des Lateinischen beschäftigten, mithin also mit genau den Sprachen, die im Testcorpus vorherrschen. Es ist also im Rahmen der Tests ein Beispielcorpus zu bilden, in dem die maximale Trefferpräzision und der maximale Trefferumfang bestimmbar ist, um die mithilfe der existierenden und zu entwickelnden Verfahren mögliche Präzision und Trefferumfang damit vergleichen zu können.

Als vorläufiges Fazit bis zur technischen Vollendung der Anwendung³⁶ kann man aber Folgendes sagen: Die Qualität einer fachspezifischen Such-

³⁵ Zum Beispiel Strunk, wie Anm. 15, oder Liar-Randell, wie Anm. 13.

³⁶ Vgl. dazu den folgenden Beitrag von Markus Heller.

maschine für Historiker beruht nicht so sehr darauf, dass die Anbieter umfangreiche Tiefenerschließung leisten und inhaltliche Normen zum Beispiel in gemeinsamen Thesauri verwenden, sondern zum einen auf einer technischen Entscheidung auf Seiten der Anbieter und zum anderen auf fachspezifischer Softwareentwicklung auch im Bereich der Suchmaschinenindices. Die Anbieter wären darauf zu verpflichten, dass die Datenstrukturen von der Suchmaschine ausgelesen werden können, das heißt dass die Informationen möglichst so in XML-Formaten auf den Servern abgelegt werden, dass die Suchmaschine Zugriff auf die XML-Rohdaten bekommt, sowie dass die Daten mit gemeinsamen XML-Vokabularien beschrieben werden. Die Fachinformatiker tragen zum Erfolg einer fachspezifischen Suchmaschine bei, indem sie die notwendigen ‚Übersetzungstools‘, wie Lemmatisierung, Graphienausgleich, Lautausgleich, Approximation, wo nötig entwickeln, auf jeden Fall testen und schließlich in die Suche integrieren.

Fachspezifische Indexierung von historischen Dokumenten II

Ein Framework zur approximativen Indexierung semistrukturierter Dokumente

Markus Heller

1. Einleitung

Die Nutzung von Rechnern in den Geschichtswissenschaften ist inzwischen weit verbreitet: Niemand schreibt seine Aufsätze mehr mit der Schreibmaschine und jeder speichert seine Entwürfe digital auf Datenträgern. Auch Quellen werden inzwischen mit Computern bearbeitet: Man bedient sich generischer Programme aus den Office-Schienen der großen Hersteller und baut auf diese Weise seine Corpora auf.

Solange derartige Editionsansätze auf ein lokales Nutzerumfeld beschränkt bleiben, mag der fehlende methodische Ansatz im Umgang mit diesen Ressourcen den geringen Einsatz rechtfertigen. Auf diese Weise ist jedoch in den vergangenen Jahren eine bunte Landschaft von hinsichtlich der Datenformate inkompatiblen und höchst heterogenen Datenbeständen gewachsen, wobei die Nutzbarkeit der Daten durch zukünftige Historikergenerationen alles andere als gesichert ist. Sind heute schon die älteren Winword-Dateiformate mit aktueller Software vom gleichen Hersteller nicht mehr lesbar, so tritt dieses Problem bei unterschiedlichen Herstellern mit proprietären Datenformaten noch in viel schärferer Form auf. Das oftmals fehlende Bewusstsein, auf offene Speicherformate und Standards zu setzen, gleicht einer fahrlässigen Strategie, öffentliches Geld und die wertvolle Lebenszeit vieler Quelleneditoren gleichsam zu vergeuden. Stattdessen liegt es nahe, wenn Editionsprojekte bei der Entscheidung über die Fi-

nanzierung maßgeblich dahingehend beurteilt werden, wie sehr sie sich der Forderung beugen, dass sie unabhängig von ihrer geographischen Speicherung und unabhängig vom zukünftigen Bearbeitungswerkzeug des Benutzers durchsucht werden können.

Dieses Szenario verlangt nach einer verteilten Sucharchitektur, in der die Daten in einem offen gelegten Format vorgehalten, zentral indexiert und mit einer ebenfalls zentralen Schnittstelle durchsuchbar gemacht werden. Dabei wird auch deutlich, dass die allgemein öffentlich verfügbare Suchtechnik von Google, Yahoo und Co. deutliche Schwächen aufweist: Sie sind auf verlinkte Webseiten unstrukturierten Inhalts spezialisiert und nehmen keinerlei Rücksicht auf Dokumente mit Auszeichnungen und hinterlegten fachlichen Zusatzinformationen. Die Merkmale solcher semistrukturierter Dokumente können bislang bei der Suche nicht angegeben werden. Auch in Bezug auf den Inhalt der Dokumente zeigen sich Probleme: Falls ein Name oder eine Bezeichnung in einem Dokument auch nur geringfügig abweichend von der Suchanfrage geschrieben wird, besteht mit der gegenwärtig öffentlich verfügbaren Suchtechnik keine Möglichkeit, diese Stellen zu finden.

Es wird also deutlich, dass ein dringender Bedarf zur Entwicklung von Suchmaschinentechologien besteht, um die Arbeit der Editoren von heute nicht wertlos erscheinen zu lassen.

2. Problemstellung

Die zu beschreibenden Probleme beginnen bei der Erfassung der Texte: Als Speicherformat bieten sich verschiedene XML-Standards an, wie etwa bei mittelalterlichen Urkunden der Standard CEI.¹ XML-Texte weisen jedoch selten eine Verlinkungsstruktur ähnlich der von HTML-Texten auf, obgleich dies grundsätzlich möglich wäre: XHTML ist eine völlig XML-konforme HTML-Variante. Ein Crawler, der alle Dokumente in einem im Inter-

¹ Adresse: <http://www.cei.lmu.de/> (letzte Einsichtnahme am 17.05.2006). Zudem: Georg Vogeler: Europäisches Urkundenerbe. Zu Potentialen und Perspektiven eines internationalen Fachinformationssystems digitaler Urkundenpublikationen. In: Elektronische Fachinformationssysteme in der Geschichte. Jahrestagung der AGE, 25.–26.11. 2005. Hg. von Franz Götz. München 2006 (im Druck).

net verteilten Corpus erschließt, kann nur jene Dokumente finden, die von anderen Dokumenten aus referenziert und verlinkt sind.

Hinsichtlich der Techniken des Information Retrieval muss zum Verständnis der Technologie erklärt werden, dass eine Suchmaschine nicht erst zum Zeitpunkt der Anfrageverarbeitung beginnt, Dokumente zu suchen. Vielmehr gleicht eine Suchmaschine einem System, das einen Index eines Buches erstellt. Der Index wird lange vor dem Suchzugriff erstellt und zum Zeitpunkt der Anfragebearbeitung nur abgefragt. Aus diesem Grund unterscheidet sich jede Suchmaschinentechologie konzeptionell von den klassischen Suchverfahren. Auch reguläre Ausdrücke dienen in erster Linie einer linearen Suche, und nicht einem Look-up in einem Index. Aus diesem Grund konzentriert sich unser Projekt nicht auf die Entwicklung von Suchverfahren, sondern auf Verfahren, in einem Index auf effiziente Weise die entsprechenden Einträge zu finden.

Bei flachen Texten gleicht nun die Dokumentenverarbeitung einer Suchmaschine der Erstellung eines Buchindizes. Doch XML-Texte weisen ein weiteres zentrales Merkmal auf, in dem sie sich von flachen Texten unterscheiden: Sie enthalten eine Struktur, anhand derer die verschiedenen Textblöcke ausgezeichnet und beschrieben werden. Mit Blick auf historische Quellentexte können mit derartigen Textstrukturen einerseits tatsächliche Eigenschaften und Gliederungsmerkmale des Originals, aber auch Hinweise eines Editors, Übersetzungen, Referenzen auf die Sekundärliteratur sowie auf andere Dokumente enthalten sein. Es liegt nun nahe, dass eine Suchmaschine zwischen all diesen verschiedenen Texttypen unterscheiden können sollte. Wer ausschließlich nach Originaltexten sucht, sollte ausschließlich Stellen in diesem Texttyp angeboten bekommen.

Interessant ist in diesem Kontext auch, dass Editoren zwar an die Einhaltung ihrer spezifischen XML-Syntaxdefinition (‚Schema‘, oder ‚Document Type Definition‘, bzw. DTD) gebunden sind, diese aber durchaus enorme Freiheitsgrade definieren kann. So kann ein XML-Schema unterschiedliche Auszeichnungstiefen zulassen. Es sind auch Rekursionen an Auszeichnungsmerkmalen möglich. Ebenfalls genannt werden muss an dieser Stelle die Tatsache, dass ein Ersteller eines Schemas die Semantik der Tags frei wählen kann. Wenn etwa ein Tag ‚document‘ definiert werden sollte, und damit gemeint ist, dass in jedem Tag ein einzelnes Quelldokument abgelegt werden darf, dann tritt die Situation ein, dass eine solche XML-Datei mehrere Quelldokumente enthält. Die Identität zwischen

‚Datei‘ und ‚Dokument‘ ist definitiv nicht mehr zwingend gegeben. Dies bedeutet, dass eine Suchmaschine für XML-Texte nicht Dateien finden will, sondern Stellen in einem XML-Corpus.

Im Allgemeinen wird von den Auszeichnungselementen in XML-Dateien von ‚Struktur‘ und von den Textinhalten von ‚Content‘ gesprochen. Wir haben nun bisher die Struktur betrachtet. Doch auch der Content weist Eigenschaften auf, welche klassische Suchmaschinen an den Rand ihrer Fähigkeiten führen. Betrachten wir die *Lebensgeschichte des Till Eulenspiegel*. Es handelt sich dabei um ein frühneuhochdeutsches Textbeispiel, das zu dieser Zeit weite Verbreitung fand und auch heute noch allgemein bekannt ist. Die elektronische Fassung eines Drucks von 1540² enthält dabei folgende Schreibungen desselben Namens:

- Ulenspiegel 288
- Ulenspiegels 6
- Ulenspiegel 2
- Ulenspiegeln 1
- Ulenspieg 1
- Ulnspiegel 708
- Ulnspiegels 11
- Ulnspiegel 1
- Ulnspiegln 1
- vlnspiegel 1
- Ulnspiegls 1
- Ulnspeigel 1
- Ulnspiegeln 1
- Ulnspiegl 1

Flexionen werden in dieser Aufzählung unterschieden, weil wir in erster Linie von einer Suchmaschine ausgehen, die keinerlei approximative Eigenschaften hat. Anhand dieses Beispiels wird die Eigenschaft von alten Texten deutlich, dass eine Orientierung an scheinbar verbindlichen Standards nicht angenommen werden kann.³ Vielmehr ist zu beobachten, dass

² Adresse: <http://www.eulenspiegel.is.guad.de/> (letzte Einsichtnahme am 17.05.2006).

³ Ob dies angesichts der häufig revidierten Vorgaben der neuen deutschen Rechtschreibung angenommen werden kann, wäre eine zu weit führende Diskussion. Das Problem selbst ist doch im Grundsatz noch präsent. Ein Ansatz für einen Indexer für eine orthographisch nicht standardisierte Sprache findet sich hier: Jan Strunk: Information retrieval for languages that lack a fixed

eine Normierung von Schreibweisen erst mit der breiten Verfügbarkeit der Wörterbücher der Gebrüder Grimm (begonnen in der ersten Hälfte des 19. Jahrhunderts) und Johann Christoph Adelung (Ende 18. Jahrhundert) festgestellt werden kann. Die Eigenschaft der Informatik, grundsätzlich von ‚sauberen‘ und normierten Daten auszugehen, führt hier nicht weiter. Auch Google unterliegt diesem Phänomen: Bei der Eingabe von ‚Schiffahrt‘ und ‚Schiffahrt‘ erhält man unterschiedliche Treffermengen. Mit Blick auf den Content von in XML-Dateien erfassten Quellentexten sind jedoch Ansätze vonnöten, die in der Lage sind, abweichende Schreibungen ebenfalls zu finden. Diese Forderung entspricht dem Postulat der deskriptiven Linguistik, die versucht, die Sprachstände abzubilden und so zu akzeptieren, wie sie in den Quellen vorgefunden werden. Eine zu Recht stark kritisierte präskriptive Linguistik würde versuchen, den Quellentexten ex post eine Norm überzustülpen, die der diachronen, aber auch der synchron-dialektologischen Variabilität der deutschen Sprache nicht gerecht wird.

Zuletzt muss sich inzwischen jede Suchmaschine an den Interessen des Benutzers orientieren: Google versucht, über Gmail Interessenprofile der Nutzer zu gewinnen, und auf diesem Weg ein Ranking anzubieten, das den Neigungen des Gmail-Benutzers entgegenkommt. Die anonyme Suchschnittstelle ist dagegen nicht personalisiert, wo gleich bei einigen Herstellern dennoch über Cookies versucht wird, mehr über die Interessenlagen der Benutzer zu erfahren.

Eine Suchmaschine für historische Texte hat im Wesentlichen drei Nutzergruppen: Historiker, Diplomatiker und Linguisten. Jede der Benutzergruppen hat ihr eigenes Interessenprofil:

- Der Historiker interessiert sich in erster Linie für Inhalte von Urkunden, aber auch für Interpretationen und Einordnungen, die von einem Editor vorgenommen wurden. Weniger interessant sind Informationen über die äußere Form von Dokumenten. Eine approximative Suche ist schon mit Blick auf verschiedene Schreibweisen von Namen unabdingbar.
- Der Diplomatiker interessiert sich weniger für den Inhalt, dafür aber für Informationen zur Erstellung und zur äußeren Form. Da modernsprachliche Anmerkungen sich eher einer orthographischen Norm unterwerfen als alte Texte, ist ein approximativer Zugriff nicht von dersel-

orthography. 2003. Adresse: <http://www.linguistics.ruhr-uni-bochum.de/~strunk/LSreport.pdf> (letzte Einsichtnahme: 21.4.2006).

ben herausragenden Bedeutung wie bei Originaltexten. Dennoch ist eine Approximation in vielen Fällen hilfreich.

- Der Linguist interessiert sich bei einem synchronen Erkenntnisziel für verschiedene Schreibungen und Aussprachen in Bezug auf eine regionale Verteilung. Bei diachronen Fragestellungen ist von Interesse, wie sich bestimmte Begriffe unter Einschränkung auf bestimmte Regionen in ihrer Form verändert haben. Dabei ist denkbar, dass einerseits approximative Fragestellungen, aber auch die Suche nach bestimmten Einzelformen von Interesse sein können. Weniger von Interesse ist für Linguisten die Suche in Annotationen.

Auf jeden Fall wird deutlich, dass bei einer Suchmaschine für semistrukturierte alte Texte der Einsatz von Approximation vom Benutzer steuerbar sein muss. Hinsichtlich der Gewichtung von Feldern sollte der Benutzer eine Möglichkeit geboten bekommen, ein seinen Interessen entsprechendes Profil auszuwählen. Die Approximation erfolgt dabei stets bezogen auf den Content. Der Strukturanteil der Suchanfrage wird dabei durch den Suchenden entweder fest vorgegeben oder in Teilen völlig freigestellt.

Neben einer direkten, einstufigen Suche besteht ein weiteres Suchschema in dem Ansatz, dass ein Benutzer nicht explizit nach einem bestimmten Begriff sucht, sondern dass er einen Sammelbegriff eingibt, um in einem zweiten Schritt seine Suche einzugrenzen. Die Funktion, die diese Möglichkeit zur Verfügung stellt, wird ‚Dynamic Drilldown‘ genannt und wird von ‚FAST Data Search‘⁴, aber auch von der Suchimplementierung von Exalead⁵ geboten. Sie setzt voraus, dass Extraktoren definiert wurden, die bestimmte Dokumenteneigenschaften oder Entitäten mit einer bestimmten Eigenschaft auszeichnen. Wie bei Exalead sind ‚Drilldowns‘ auf Eigennamen, Ortsnamen, Regionen, Sprachen, Dokumententypen, Archive usw. möglich. Ein derartiger Zugriff würde der Fragestellung entgegenkommen, nach der ein Historiker etwa mehr über das geographische oder personelle Umfeld einer Person oder eines Vorgangs wissen will.

Das bislang bezeichnete Szenario ist sehr umfangreich und im Rahmen eines kleinen Projekts nur äußerst bedingt umsetzbar. Mitunter aus diesem Grund haben wir uns entschlossen, nur die wesentlichen Technologiefragen zu klären und Themenbereiche auszuklammern, von denen wir über-

⁴ Adresse: <http://www.fastsearch.no> (letzte Einsichtnahme am 17.05.2006).

⁵ Adresse: <http://www.exalead.com> (letzte Einsichtnahme am 17.05.2006).

zeugt sind, dass sie weit über den fachlichen Horizont der Computerlinguistik bzw. Historischen Fachinformatik hinausführen. Darunter fällt beispielsweise die gesamte Fragestellung um die Erschließung verteilter Corpora: Wir gehen davon aus, dass die Texte bereits vorliegen. Ein weiterer Bereich, der besonders von Usability-Spezialisten gelöst werden muss, ist jener des Benutzerfrontends. Da die Firma Arpa Data GmbH⁶ Interesse an einer Softwarebibliothek hat, die in die bestehende OCR-Lösung integriert werden kann, wollen wir uns darauf beschränken, ein objektorientiertes Framework zu entwickeln, mit dem die bezeichneten Fragestellungen des Information Retrieval grundsätzlich und für unsere Belange exemplarisch gelöst werden können.

3. Projektbeschreibung

Das beschriebene Projekt hat sich zum Ziel gesetzt, eine Indexierungsarchitektur zu entwickeln, die gemischte, also Content- und Strukturanfragen (,CAS-Queries') beantworten kann, wobei der Content-Anteil über Approximationsverfahren im Zugriff auf den Index aufgeweicht werden soll. Auf diese Weise können Anfragen bearbeitet werden, deren Strukturanteil sich auf die Auszeichnung durch einen Fachmann bezieht, und deren Content-Anteil mit möglichen abweichenden Schreibungen in Deckung gebracht werden kann.

Es gibt bereits seit einigen Jahren Datenbanksysteme, die nativ mit XML-Texten umgehen und CAS-Abfragen beantworten können.⁷ Allerdings ist zu beobachten, dass bis dato keine XML-Datenbanken oder native XML-Indizes bekannt sind, die eine Approximation auf den Content- oder Strukturanteil implementieren. Unsere Entscheidung, einen eigenen Index zu entwickeln, lag darin begründet, dass der Aufwand, eine etablierte Matching-Engine⁸ einer Datenbank auf approximative Fähigkeiten hin zu erweitern, im Arbeitsumfang schwer kalkulierbar ist. Es besteht das nicht un-

⁶ Adresse: <http://www.arpa.ch> (letzte Einsichtnahme am 17.05.2006). Arpa Data beschäftigt sich vor allem mit der OCR-technischen Erfassung von Zeitungsarchiven.

⁷ Vgl. Harald Schöning: Tamino – A Database System Combining Text Retrieval and XML. In: Intelligent Search on XML Data. Applications, languages, models, implementations, and benchmarks. Ed. by Henk Blanken. Berlin 2003. S. 77–94. Auch: Cynthia M. Saracco: Query DB2 XML Data with SQL 16.03.2006. Adresse: <http://www-128.ibm.com/developerworks/db2/library/techarticle/dm-0603saracco2/> (letzte Einsichtnahme 06.04.2006).

erhebliche Risiko, dass ein neues Release aus dem Hause des Herstellers die hinzugefügten Approximationsfunktionen nicht unterstützt oder die Modifikationen inkompatibel sind. Aus diesem Grund haben wir uns zu einer Eigenentwicklung entschlossen, zumal ein großer Teil der benötigten Teilalgorithmen an unserem Institut oder in Kooperation entwickelt wurden. Durch die Möglichkeit, die Rahmenbedingungen der Architektur selbst zu definieren, konnten wir auch die Anforderungen aus dem Retrieval in historischen Textcorpora einfließen lassen.

In Bezug auf historische Texte ist erkennbar, dass die Contentapproximation nicht nur alphabetisch, sondern auch gesondert numerisch erfolgen muss, wenn etwa nach Dokumenten um das Jahr 1400 gefragt wird: Es ist neben dem alphabetischen ein weiterer Lookup für Zahlen und Jahresangaben erforderlich. Über eine konfigurationsgesteuerte Anwahl eines bestimmten Indextyps werden als Zahlen erkannte bzw. in Zahlen konvertierte Entitäten typenorientiert geroutet. Die Unterscheidung ist aus dem Grund wichtig, weil sich bei einer alphabetischen Approximation die Fehlerschwere zwischen den verschiedenen Stellen eines Wortes kaum unterscheidet. Bei Zahlen ist die Fehlerschwere aber sehr wohl an die Stelle gebunden, denn der Unterschied zwischen „1991“ und „1992“ ist eher gering und zwischen „1004“ und „2004“ doch enorm. Außerdem ist die stelleninhärente Approximation ebenfalls eine andere: Wenn man annimmt, dass ein falscher Buchstabe mit einem Fehlermaß des Wertes 1 belegt wird, so ist das Wort womöglich nicht mehr im Wortschatz der Sprache enthalten und ‚falsch‘. Eine Jahreszahl wird anders, hingegen nie falsch sein.

Es ist durchaus denkbar, neben dem Wort- und dem numerischen Index einen dritten Typ zu entwickeln: Ein geographischer Index könnte auf Koordinaten Bezug nehmen und eine geographische Näherung bereitstellen. Man könnte also eine Suche auch auf Dokumente mit einem Bezug auf gewisse Orte eingrenzen. Gerade im Bereich der Wirtschafts- und Sozialgeschichte erscheinen solche Anwendungsgebiete sehr viel versprechend, berücksichtigt man doch, dass in der letzten Zeit sehr viele Ansätze über GIS-Anwendungen vorgestellt worden sind. Besonders zur *XVI International Conference of the Association for History and Computing 2005* in Amsterdam

⁸ Erst jüngst ist bekannt geworden, dass der Datenbankhersteller MySQL eine Schnittstelle für die Einbindung von Storage Engines von Drittanbietern definieren will: Datenbank MySQL will mehr Storage-Engines einbinden. Heise-News, 26.04.2006. Adresse: <http://www.heise.de/newsticker/meldung/72417> (letzte Einsichtnahme am 17.05.2006).

konnten viele Projektberichte zu GIS-Themen beobachtet werden.⁹ Allerdings wird im Rahmen des Projekts eine derartige Erweiterung aus Kapazitätsgründen nicht verfolgt.

Im Bereich der Stringologie wird das hier offenkundige Problem ‚the nearest neighbour problem‘ genannt. Allerdings ist nicht nur der nächstbeste Treffer gesucht, sondern alle Stellen im Corpus, die in bestimmten Editierdistanz-Grenzen die in der Anfrage genannten Anforderungen erfüllen. Die Approximation darf hier nicht wie in den meisten der bislang in der Literatur beschriebenen Approximationsverfahren auf einer linear vergleichenden Suche aufbauen.¹⁰

Das Ergebnis ist meist eine längere Ergebnisliste, wobei die Errechnung des Rankings in zwei Stufen erfolgt: Die erste Sortierung der Dokumente erfolgt dokumentenspezifisch, wobei bereits zum Zeitpunkt der Befüllung im Index anhand von Berechnungen zur Relevanz eines Terms in Bezug auf den Container, das Gesamtdokument und den Gesamtkorpus mittels TF/IDF-Verfahren festgelegt wird, wie relevant es in Bezug auf andere Container ist.

Es entsteht also bereits zum Zeitpunkt der Befüllung („Feeding“) ein großes Ranking innerhalb kriterien gleicher Ablagen auf der Platte. In einem zweiten Schritt, der erst zum Zeitpunkt der Anfragebearbeitung erfolgt, wird anhand eines Gewichtungspfils von Containertypen errechnet, wie die Treffermenge sortiert werden muss. Das Gewichtungsprofil entspricht den Anforderungen der unterschiedlichen Interessenlagen der Benutzer. Hier werden die Interessen des Historikers, des Linguisten oder des Diplomaten abgebildet. Die bezeichneten Profile sind jederzeit änderbar, ohne dass der Index neu erstellt werden muss.

Da vorab nicht bekannt sein kann, für welche Dokumente die Sucharchitektur als solche eingesetzt werden wird, haben wir uns entschieden, dem Beispiel der FAST-Suchmaschine folgend einen modularen Verarbeitungsansatz zu wählen. Es ist dem Betreiber überlassen, die Dokumentenverarbeitung so zu konfigurieren, wie es den Eigenschaften des Corpus entspricht. Besonders die Informationsextraktion, also die automatische Iden-

⁹ Vgl. Royal Netherlands Academy for Arts and Sciences. Humanities, Computers and Cultural Heritage. Proceedings of the XVI International Conference of the Association of History and Computing. 14–17 September 2005, Amsterdam. Adresse: <http://www.knaw.nl/publicaties/pdf/20051064.pdf> (letzte Einsichtnahme am 17.05.2006).

¹⁰ Zu linearen String verarbeitenden Algorithmen: Maxime Crochemore und Wojciech Ritter: *Jewels of Stringology*. New Jersey 2003. Darin besonders S. 183–198.

tifikation von Entitäten, kann mit Modulen erfolgen, die je nach Bedarf konfiguriert oder gar vom Betreiber selbst programmiert sein können. Die Dokumentenverarbeitung umfasst dabei üblicherweise Schritte wie die Feststellung des Encodings des zu verarbeitenden Dokuments, eine Konvertierung in Unicode/UTF-8, die Tokenisierung und, sofern gewünscht, eine Identifikation von weiteren Strukturmerkmalen wie Satzgrenzen oder Entitäten, wie sie durch lokale Grammatiken erkennbar sind.

Im Rahmen der Pipeline wird auch festgestellt, ob eine Entität nicht im alphabetischen Standardindex, sondern im numerischen oder gegebenenfalls im zukünftig optionalen Geo-Index abgelegt werden soll. Die Typisierung der Entität erfolgt in jedem Fall während der Dokumentenverarbeitung und damit außerhalb der eigentlichen Indexstruktur. Damit bleibt der Indexer selbst schlank und performant.

Der Indexer unterscheidet zwischen den jeweiligen getypten Content- und den Strukturanteilen. Für Strukturanteile wird eine fortlaufende Nummer vergeben, wobei diese bei einer fehlenden Strukturinformation, also bei flachen unstrukturierten Dokumenten, stets null ist. Das bedeutet, dass die Indexerarchitektur flache und damit nicht ausgezeichnete Dokumente ebenfalls verarbeiten kann. Sofern die Dokumentenverarbeitung ohne eine vorliegende Strukturinformation trotzdem andere Indextypen, also Zahlen oder geographische Informationen, erkennt, so wird die Indexerarchitektur sie problemlos indexieren.

In Bezug auf die historisch-fachinformatische Aufgabenstellung kann darauf hingewiesen werden, dass jegliche anwendungsspezifischen Parameter aus der Konfiguration der Dokumentenverarbeitung sowie der Abfrageverarbeitung resultieren. Die Architektur des Indexers selbst ist generisch und flexibel genug, um vielen Bedürfnissen des modernen Information Retrieval zu genügen.

Es wurde beim Design der Indexarchitektur versucht, ein Optimum an Zugriffsgeschwindigkeit und Skalierbarkeit zu finden. Dies ist durch Verwendung der Methoden von Felix Weigel gelungen, der den ‚ContentAware DataGuide‘ entwickelt hat. Anders als in seiner Referenzimplementierung haben wir auf eine Datenbank verzichtet und verwenden für die Speicherung der Referenzen auf den Corpus reguläre Dateien und zur Verwaltung dieser die Mechanismen des Dateisystems. Auf diese Weise sparen wir Code zur Verwaltung der Referenzen und die Architektur gewinnt deutlich an Stabilität. Das Maximalmaß der Skalierbarkeit einer Implementierung

liegt bei den Grenzen der Partition, in welcher der Index gespeichert wird. Dabei kann es sich auch um ein logisches Volume handeln, das über mehrere physikalische Volumes verteilt ist. Mit derartigen Verfahren liegt die theoretische maximale Skalierbarkeit des Indexers im unteren Terabyte-Bereich. Tests stehen im Moment noch aus, sind jedoch bereits geplant. Als Testdaten sollen Genom-Sequenzierungsdaten dienen, die in den entsprechenden Größenordnungen als XML-Dateien vorliegen.

Alle Kernkomponenten des Indexers sind streng in C++ und die Einbindung von Modulen ist in Python gehalten. Die in Entwicklung befindliche Suchmaschine instantiiert daher nur verschiedenste Klassen einer Bibliothek und führt sie in einer Applikation zusammen. Auf diesem Weg ist es möglich, die Klassen als Bibliothek gesondert abzulegen: unter Unix als ‚Shared Object‘ und unter Windows als DLL. Die Entwicklungsplattform ist jedoch Linux/Unix.

Von den großen, kommerziellen Vorbildern unterscheidet sich unser Projekt vor allem dadurch, dass Elemente fehlen, die für eine kommerzielle Infrastruktur typisch sind: Wir beabsichtigen, uns auf die Fragestellungen aus der Indexierung zu konzentrieren und überlassen die Implementierung von proaktiven Monitoringsystemen sowie von Ansätzen der verteilten bzw. clusterorientierten Architektur den Benutzern der Bibliothek, die derartige Eigenschaften benötigen.

4. Architekturdetails

Das Projekt hat sich zum Ziel gesetzt, eine Indexierungsarchitektur zu entwickeln, die Content-and-Structure(CAS)-Anfragen verarbeiten kann, wobei der Content-Anteil approximativ und getypt verarbeitet werden kann. Die Architektur soll modular, flexibel, skalierbar, robust und portierbar sein. Ein großer Teil der verwendeten und eingebundenen Technologien entstammen unserem Institut, die Architektur der Dokumentenverarbeitung ist an jene der Altavista- und der ‚FAST-DataSearch‘-Engine angelehnt.

Im klassischen XML-Retrieval wird jeweils der Strukturanteil einer Anfrage einerseits und der Contentanteil der Anfrage andererseits ausgewertet. Zur Laufzeit der Anfragebearbeitung wird ein Join der beiden Ergebnismengen berechnet, wobei die beiden Ergebnismengen erst vom Plattenspeicher in den Arbeitsspeicher geladen werden müssen. Felix Weigel, auf dessen An-

sätzen wir unsere Architektur maßgeblich begründen, kritisiert zurecht, dass beim Laden der Ergebnismengen unnötig viele ‚false positives‘ geladen werden müssen. Außerdem ist richtig, dass die Berechnung des Joins zur Laufzeit der Anfragebearbeitung die Ausgabe des Ergebnisses massiv und unnötig verzögert. Sein Konzept eines ‚Content-Aware DataGuide‘ nimmt die Berechnung des Joins vorweg und bietet über einen Government- bzw. Containment-Test die Möglichkeit, die Erfüllbarkeit der Anfrage zu prüfen, noch bevor von der Platte Locations geladen werden.¹¹

Allerdings überlässt die Referenzimplementierung Felix Weigels die Speicherung der Daten über eine JDBC-Schnittstelle einer Datenbank. Wie oben bereits kurz ausgeführt, sehen wir in diesem Ansatz drei deutliche Nachteile: Falls eine JDBC-Implementierung verwendet wird, die auf einem Netzwerk-Socket basiert, so wird die Zugriffsgeschwindigkeit durch die Abarbeitung der Netzwerkroutrinen unnötig verzögert. Dies ist auch der Fall, wenn über das Loopback-Interface auf den Localhost zugegriffen wird. Hier entfällt natürlich die sonst zudem einzuberechnende Latenz bei Netzwerkverbindungen.¹² Der zweite Nachteil liegt darin, dass die Matchingfunktionen der Datenbank überlassen bleiben. Eine Approximation über die Möglichkeiten des LIKE-Operators hinaus ist außerhalb des SQL92-Standards kaum denkbar, will man sich nicht an einen bestimmten Datenbankhersteller binden. Es ist zwar richtig, dass verschiedenste Datenbankhersteller Operatoren entwickelt haben, deren Flexibilität weit jenseits derer des LIKE-Operators liegen. Aber die Modifikation der Matching-Engine einer Datenbank insgesamt erschien uns kaum kalkulierbar. Als dritten Nachteil sehen wir, dass die Verwendung der SQL-Schnittstelle insgesamt unnötig ist. Datenbanken sind generische Werkzeuge zur Speicherung von Daten. Auch wenn die Datenbanktechnologie sehr weit fortgeschritten ist und enorm leistungsfähige Verfahren zur Verwaltung von Daten hervorgebracht hat, so bindet einerseits die Formulierung, andererseits das Parsing und die Verarbeitung von SQL-Ausdrücken Prozessorzeit. Wir sind der Überzeugung, dass betriebssystemnähere und problemorien-

¹¹ Felix Weigel, Holger Meuss, François Bry und Klaus Schulz: Content-Aware DataGuides: Interleaving IR and DB Indexing Techniques for Efficient Retrieval of Textual XML Data. In: Advances in information retrieval. Proceedings of the 26th European Conference on Information Retrieval (ECIR). Ed. by Sharon McDonald und andere. Berlin, Heidelberg 2004. S. 378–393.

¹² Adresse: <http://java.sun.com/products/jdbc/> (letzte Einsichtnahme am 17.05.2006). Offizielle JDBC-Homepage der Firma SUN Microsystems.

tierte Ablagestrukturen gegenüber dem Einsatz von Datenbanktechnologien deutliche Geschwindigkeitsvorteile bieten.

In der Implementierung der Ansätze aus dem CADG unter Verzicht auf eine relationale Datenbank und mit einer eigenen Verwaltung der Content-Tokens sehen wir die Möglichkeit, eine weitere Technologie aus dem Umfeld unseres Instituts zum Einsatz zu bringen: Stoyan Mihov und Klaus Schulz haben ein Verfahren vorgestellt, mit dem in einem Wörterbuch-Automaten sehr schnell approximativ gesucht werden kann.¹³

Die Flexibilität der Architektur wird dadurch sichergestellt, dass der größte Teil der Vorverarbeitung der Dokumente durch Module erfolgt, die der Betreiber der Infrastruktur nach seinen Bedürfnissen konfigurieren kann. Die Schnittstelle zu den Modulen ist genauso wie bei ‚FAST Data-Search‘ in der Scriptsprache Python gehalten.

Die Relevanzsortierung ist ein Thema, das sich wie natürlich an die Fragestellung der Indexarchitektur anschließt. So verwundert es nicht, dass sich der Autor des ‚Content-Aware DataGuide‘, Felix Weigel, ebenfalls mit dieser Frage befasst hat.¹⁴ Der Inhalt seiner Überlegungen betrifft die Frage, wie verschiedene Rankingmodelle mit dem Verfahren des vorberechneten Joins zwischen Struktur- und Contentinformation integriert werden können. Seine Betrachtungen beziehen sich dabei auf XPRES¹⁵, BUS¹⁶, XIRQL¹⁷ und XXL¹⁸.

¹³ Stoyan Mihov und Klaus Schulz: Fast Approximate Search in Large Dictionaries. In: Computational Linguistics 30 (2004). S. 451–477.

¹⁴ Felix Weigel, Holger Meuss, Klaus Schulz und Francois Bry: Content and Structure in Indexing and Ranking XML. In: Proceedings of the 7th International Workshop on the Web and Databases (WebDB 2004). 17–18 June 2004, Paris, France. Ed. by Sihem Amer-Yahia und Luis Gravano. O.O. O. J. S. 67–72..

¹⁵ Jens E. Wolff, Holger Flörke und Armin B. Cremers: Searching and Browsing Collections of Structural Information. In: Proceedings of the IEEE Forum on Research and Technology Advances in Digital Libraries. Washington 2000. S. 141–150.

¹⁶ Dongwook Shin, Hyuncheol Jang und Honglan Jin: BUS. An Effective Indexing and Retrieval Scheme in Structured Documents. In: Proceedings of the 3rd International Conference on Digital Libraries. Pittsburgh 1998. S. 235–243.

¹⁷ Norbert Fuhr und Kai Großjohann: XIRQL: A Query Language for IR in XML Documents. In: Proceedings of the 24th Annual International Conference on Research and development in Information Retrieval. ACM. New York 2001. S. 172–180.

¹⁸ Anja Theobald und Gerhard Weikum: The Index-based XXL Search Engine for Querying XML Data with Relevance Ranking. In: Proceedings of the 8th International Conference on Extending DB Technology. Prag 2002. S. 477–495.

Die atomare Einheit in unserem Index und damit das Objekt, das der Benutzer zu finden gedenkt, ist ein spezifischer Token in einem spezifischen Container. Eigentlich soll in der Regel eine Location gefunden werden, also eine Stelle in einem Text. In Bezug auf semistrukturierte Dokumente kann jedoch nicht mehr von der semantischen Homogenität eines Dokuments ausgegangen werden, und so nimmt der XML-Container mit einem bestimmten Pfad die Stelle ein, die in der Bewertung flacher Corpora das Dokument hat. Es soll also nach einer Stelle in einem bestimmten Container gesucht werden. Da die Stelle jedoch aus einer Kombination von Dokument und XML-Pfad besteht, muss jede Stelle eine Nummer enthalten, damit nicht nur zwischen den Dokumenten im Corpus, sondern zwischen allen Containern unterschieden werden kann.

Die Anfrage sucht stets nach den Containern, in denen ein bestimmter Content eine herausragende, bezeichnende Bedeutung hat. Insofern muss beim Befüllen des Index und beim Erstellen der Rankingparameter bedacht werden, ob ein bestimmter Token in einem bestimmten Container relevant ist. Dies kann durch die Berechnung der ‚Term Frequency‘ erfolgen. Auf diese Weise ist es möglich, je Dokument die Reihenfolge der relevantesten Wörter festzulegen.

Allerdings kann erst dann eine wirksame Aussage über die Relevanz eines Wortes in Bezug auf den Text getroffen werden, wenn die Termfrequenz zur Häufigkeit des Wortes im Corpus in Bezug gesetzt wird. Diese Bezugsetzung erfolgt rechnerisch über eine Invertierung, der ‚Inverse Document Frequency‘.

In Bezug auf semistrukturierte Dokumente muss daher die Relevanz immer in Bezug auf denselben Containertyp ermittelt werden, da es sich bei einem semistrukturierten Corpus genau genommen nicht um einen homogenen Pool an Texten handelt, sondern um eine Menge sehr spezifisch ausgezeichnete Texte, die in den jeweiligen Containern enthalten sind. Aus diesem Grund modifizieren wir die Formel ‚TF/IDF‘ zu ‚TF/ICF‘ (Inverse Container Frequency).

Der Wert, der sich aus der Berechnung von TF und ICF ergibt, ist das Sortierungskriterium der Location-Liste in den LL-Dateien. Er wird zusammen mit den Location-IDs abgespeichert, um eine Einfügung an einer entsprechenden Stelle zu erlauben, wenn weitere Daten im Index abgelegt werden.

Für die Berechnung von TF/ICF es ist nötig, die Containerfrequenz des jeweiligen Wortes zu kennen. Daher steht sie an der ersten Stelle der Location-Liste. Beim Einfügen eines neuen Eintrags muss die Location-Liste ohnehin geladen werden, und so kann ohne großen gesonderten Aufwand die Information über die ICF des jeweiligen Wortes gleich in den Speicher geladen und zur Verarbeitung bereitgestellt werden.

In Bezug auf die von Felix Weigel betrachteten Rankingverfahren lehnt sich unser Verfahren an BUS an. Allerdings trennen wir klar zwischen der Relevanzfeststellung des Contentanteils und des Strukturanteils, da letzteres in unserer Gesamtarchitektur eine besondere Rolle spielt: Die Bewertung des Strukturanteils fließt zum Zeitpunkt der Befüllung noch nicht mit ein, sondern über ihn bilden wir die Präferenzen der verschiedenen Benutzergruppen ab. Doch dazu später mehr im Abschnitt über das Ranking.

Die Komponenten der Architektur sind in einzelnen:

- der Feeder, welcher folgende Komponenten enthält:
 - den Parser,
 - die Document Processing Pipeline,
- der Indexer (apCADG), welcher folgende Komponenten enthält:
 - den DataGuide,
 - den WordIndex,
 - den NumericIndex,
 - als Option den Geo-Index,
 - das Lametta,
 - die StorageArea auf der Platte,
 - den QueryProcessor,
- der QueryHandler, der folgende Komponenten enthält:
 - die Query Processing Pipeline,
 - den Kontrollautomaten-Compiler,
 - den Ranker.

Die wichtigsten der genannten Komponenten sollten folgend kurz vorgestellt werden:

4.1 Der Feeder

Der Feeder ist eine Wrapper-Klasse, die das einzulesende Dokument entgegennimmt und dem SAX-Parser übergibt. Dieser zerlegt es in Pfad-Con-

tent-Tupel und übergibt jeden der Tupel an die Document Processing Pipeline. Diese implementiert eine in den C++-Code eingebettete Python-Instanz, wobei die Tupel an die Python-Instanz zur Verarbeitung übergeben werden. Die Instanz lädt je nach Konfiguration ein Framework an Python-Modulen, deren wichtigste Aufgaben folgende sind:

- Encoding-Normalisierung auf UTF-8,
- Tokenisierung,
- Annotation von Entitäten.

Es sind zahlreiche weitere Bearbeitungsstufen denkbar wie eine Language-Detection, ein Lemmatizer, ein Vectorizer, eine Date/Time-Extraction und viele mehr. Wir beschränken uns auf letztere, da sie wirklich essentiell sind. Die Abwendung von der Entität der XML-Datei als solcher ist dadurch gerechtfertigt, dass einerseits eine XML-Datei mehrere Quelltexte enthalten kann, und der Benutzer andererseits nicht daran interessiert ist, Dateien, sondern Dokumente, respektive Textstellen, im Corpus zu finden.

4.2 Der DataGuide

Man kann ein XML-Dokument als Baum darstellen, bei dem der äußerste XML-Knoten der Wurzelknoten und jeder innere Container ein entsprechender Kind-Knoten ist.¹⁹ Wenn man die Wurzelknoten aller Dokumente im Corpus zusammenfasst, und dabei alle namensgleichen Container konsolidiert, so erhält man einen Baum an all den XML-Pfaden, die im Corpus enthalten sind. Ein Baum dieses Typs wird DataGuide genannt. Jeder Knoten dieses Baumes erhält dabei eine ganzzahlige, fortlaufende Nummer, ungeachtet dessen, ob es sich dabei um einen Knoten des Schema-Typs PC-DATA (Knoten ohne weitere Kinder) handelt oder nicht.

Unsere Implementierung verwendet eine modifizierte Fassung der ‚core::tree‘-Template-Klasse von Justin Gottschlich²⁰, die insofern verändert wurde, als dass die Suche eines bestimmten Kindknotens im Baum nicht mehr linear, sondern mit einer STL-Map logarithmisch erfolgt. Wir haben

¹⁹ Zur Baumdarstellung von XML-Dokumenten siehe: Jamens Clark und Steven deRose: XML Path Language (XPath) 1.0. W3C Recommendation 16 November 1999. Adresse: <http://www.w3c.org/TR/xpath> (letzte Einsichtnahme am 17.05.2006).

²⁰ Vgl. Justin Gottschlich: C++ Trees. Adresse: <http://www.gamedev.net/reference/programing/features/coretree1/> (letzte Einsichtnahme am 17.05.2006).

das C++-Template so verändert, dass jeder Elternknoten seine Kindknoten anhand eines eindeutigen Namens identifiziert. Der Name ist dabei der Tag des entsprechenden XML-Containers.

Besondere Beachtung verdienen Attribute, denn sie kommen unter Umständen mehrfach in demselben XML-Tag vor. Wir behandeln Attribute wie Pfaderweiterungen, wobei der Inhalt des Containers bei mehreren Attributen mehrfach indexiert wird. Zudem wird die Gesamtheit aller Attribute zusammengefasst und Container auf diese Weise anhand ihrer gesamten Strukturidentifikation auffindbar gemacht.

4.3 Der WordIndex und Kompilierungsverfahren

Der Zugriff auf den Wortindex implementiert ein Verfahren nach Klaus Schulz und Stoyan Mihov,²¹ nach dem mittels Backtracking in einem Wortindex einerseits und einem Levenshtein-Kontrollautomaten andererseits bei kurzen Wörtern und einer niedrigen Editierdistanz drastische Geschwindigkeitsgewinne erreicht werden können. Das Verfahren setzt also einen bereits fertig kompilierten Wortindex voraus, dessen Kompilierung folgend kurz beschrieben werden soll.

In seiner logischen Darstellung besteht der Index aus einem klassischen Trie.²² Von einer Wurzel ausgehend führen deterministische Übergänge zu neuen Zuständen, wobei bei jeder tieferen Übergangsstufe der Präfix des eingegebenen Wortes weiter spezifiziert wird. Mathematisch wird eine derartige Struktur als ein deterministischer, endlicher azyklischer (gerichteter) Graph bezeichnet.

Würde man nun jeden Zustand des Graphen im Speicher eigens allozieren, so bräuchte man ohne Rücksicht auf die im jeweiligen Zustand real existierende Anzahl an Übergängen eine linear große Speichermenge. Da jedoch die Anzahl der Wörter eine der stark skalierenden Größen im Index ist, wird man bestrebt sein, gerade hier den Speicherverbrauch einzudämmen. Aus diesem Grund liegt es nahe, eine Kompression der Übergangstabelle nach einem Verfahren von Tarjan vorzunehmen.²³ Dieses Verfahren

²¹ Vgl. Anm. 13.

²² Zu Tries siehe: Donald E. Knuth: *The Art of Computer Programming*. Vol. 3, 2nd ed. Boston 1998. S. 492–512.

²³ Robert Endre Tarjan und Andrew Chi-Chih Yao: *Storing a sparse table*. In: *Communications of the ACM* 22 (1979). S. 606–611.

verschränkt Übergänge in einer Weise, in der die Performanz im Zugriff auf die Zustände nicht beeinträchtigt wird, wobei aber durch die Verschränkung enorm an Platz gespart wird.

In unserem Ansatz verwenden wir das Tarjan-Verfahren, jedoch erweitern wir es zu einem Transduktor, der nicht nur Eingabewörter erkennt, sondern bei einem Endzustand einen Ausgabewert zurückgibt. Der Ausgabewert in unserem WordIndex besteht aus einer Sprungadresse im Speicher und einer fortlaufenden Zahl. Die Sprungadresse zeigt auf den entsprechenden Bitstring in der Lametta-Datenstruktur. Die fortlaufende Zahl wird für die Adressierung der jeweiligen Location-Datei auf der Platte benötigt.

Der Kompilierungslauf ist in diesem Sinne eigentlich ein Kompressionslauf, der den Stack und die Liste der bereits verarbeiteten Wörter zusammenführt und sortiert und anschließend daraus den Trie in seiner kompakten Form berechnet und im Speicher zur Verfügung stellt. Dieselben Verfahren kommen bei der Kompilierung der Lamettatabelle zum Einsatz.

Der Tarjan-Mechanismus allokiert also je Zeile der Übergangstabelle entsprechend der Länge des Alphabets in einem vordefinierten und mit Null-Pointern initialisierten Speicherbereich $2n+1$ Zellen, wobei n die Mächtigkeit des Alphabets ist. Eine Zelle beinhaltet zwei Zeiger, wobei der eine auf den nächsten Zustand und der andere auf den Anfang des aktuellen Zustands zeigt, um sicherzustellen, dass der aktuelle Übergang auch ein Übergang des aktuellen Zustands ist. In Erweiterung zum Tarjan-Schema fügen wir dem Algorithmus noch einen weiteren Zeiger hinzu: Am Anfang der Tarjan-Zeile befindet sich in unserer Implementierung ein Zeiger auf den Anfang des vorhergehenden Zustands, welcher benötigt wird, um mit einer möglichst geringen Speichersignatur eine Rückwärts-Auflösung des Tokens zu ermöglichen.

4.4 Der NumericIndex

Der Index zur numerischen Approximation bewegt sich lediglich in einer einzigen, zeitlichen Dimension. Eine Approximation erfolgt daher ausschließlich numerisch. Die am besten geeignete Datenstruktur hierfür ist eine Double Ended Queue: Eine zweifach verlinkte Kette der C++ Standard Template Library, die einen Zugriff über einen Index erlaubt, aber auch Einfüge-Operationen zur Verfügung stellt. Jedes Glied der Kette ist dabei vergleichbar mit einem Endzustand und beinhaltet daher einen Zeiger auf

die entsprechende Lametta-Zeile sowie eine Zahl als Indexmerkmal für die Datenstruktur auf der Platte. Diese Zahlen sind über alle Index-Systeme hinweg fortlaufend.

Zum gegenwärtigen Stand sehen wir den Einsatz eines Konstrukts aus der STL als den effektivsten Weg, zu einer funktionierenden Architektur zu gelangen, und wollen daher nicht ausschließen, dass hier noch Optimierungspotential besteht.

4.5 Das Lametta

Das Lametta hat die folgende Funktion: Die Nummer des XML-Knotens, die der DataGuide bei der Queryverarbeitung liefert, dient als Index für eine Bitfolge, bei der jeweils das erste Bit den ‚Containment-Test‘, und das jeweils zweite Bit den ‚Government-Test‘ ermöglicht. Der Containment-Test beantwortet die Frage, ob ein XML-Knoten selbst das gesuchte Wort enthält. Der Government-Test beantwortet die Frage, ob eines der Kinder des aktuell betrachteten Knotens das Wort enthält. Der Containment-Test gibt die qualifizierte Antwort, ob es gerechtfertigt ist, mit der Nummer des DataGuide-Knotens und der Nummer des Worts aus dem Wortindex auf der Platte aus der ‚Ablagestruktur nach Contentinformation‘ die Liste der zutreffenden Locations zu holen.

Der Government-Test ist besonders bei Anfragen wichtig, bei denen sich in der Pfadinformation ein Freiheitsgrad im Sinne eines Kleene-Sternchens befindet. Hier muss der Pfadbaum in der Tiefe durchlaufen werden, um nicht Pfade zu selektieren, bei denen das gesuchte Wort ohnehin nicht enthalten ist. Wir verwenden den binären Government-Test um zu klären, ob sich ein Hinabsteigen in einen bestimmten Pfad überhaupt lohnt. Die konzeptionelle Idee hierfür entstammt den Arbeiten von Felix Weigel.

Unsere Implementierung ohne Rückgriff auf eine Datenbank arbeitet mit festen Bitstring-Längen, wobei im Gegensatz zur Kompilierung von Automaten keine Übergänge, sondern lediglich eine Bool'sche Wahr- oder Falsch-Information kodiert werden soll. Beide Informationen werden durch Gerichtetheit des Speicherwertes auf den Anfang oder das Ende der festen Zeilenlänge der Tabelle kodiert. Auf diese Weise lässt sich eine ebenso leistungsfähige, Speicher sparende Kompression wie bei der Kompilierung des WordIndex erreichen.

Sollte bei einer Neukompilierung die feste Länge des Bitstrings nicht ausreichen, so wird diese um einen festen Wert erweitert. Dieses Verfahren belegt angesichts der Tarjan-Kompressionsmethode keinen unnötigen Speicherplatz, da unbenutzte Speicherzellen weiterhin von anderen passenden Bitstrings benutzt werden können.

Die Abarbeitung der beiden Tests erfolgt, indem aus dem WordIndex auf den Speicherplatz des entsprechenden Lametta-Strings gesprungen wird. Mit der Indexzahl aus dem DataGuide wird die Position der jeweiligen Speicherzelle ermittelt. Wenn der Zeiger in dieser Zelle mit dem Zeiger auf die Lamettafolge, also die erste Speicherzelle der Zeile, identisch ist, so ist der Test positiv. Zeigt der Zeiger in der Speicherzelle auf das Ende der Lamettafolge, so ist der Test negativ. Nicht beschriebene Zellen enthalten den Null-Wert.

4.6 Die StorageArea

Zur Ablage der Locations auf der Festplatte haben wir folgende Systematik erarbeitet: Um den Durchgriff auf die Daten bei der Suche möglichst robust, schnell, skalier- und portierbar zu halten, haben wir uns an den Möglichkeiten orientiert, die gängige Betriebssysteme zur Verfügung stellen. Die Ansprache einer Locations-Datei erfolgt daher über einen Dateipfad, der nach der Abarbeitung der im Speicher befindlichen Datenstrukturen zusammengestellt wurde. Die fortlaufenden Nummern aus dem DataGuide und den Indexstrukturen werden in ihre hexadezimale Repräsentation konvertiert und in jeweils zwei Zeichen gestückelt. Anschließend wird eine Datei mit dem Namen „LL“ hinzugefügt, sodass ein Pfad wie folgt aussieht:

Aus einer DataGuide-Nummer „3f4d“ und einer Wordindex-Nummer „19bc3d“ wird nun der Pfad „<Stamm>/3f/4d/19/bc/3d/LL“. „LL“ ist dabei eine Datei, die die Liste aller Locations enthält. Der Dateiname wurde aus dem Grund auf „LL“, bzw. auf „ZZ“ gesetzt, weil diese Buchstaben nicht Teil des hexadezimalen Alphabets sind.

Bei einer hexadezimal dreistelligen Wortadressierung liegt die maximale Skalierung bei 16 777 216 möglichen Wörtern. Falls diese Anzahl übertroffen wird, so wird der bestehende Baum in ein neu zu schaffendes Unterverzeichnis „<Stamm>/00“ kopiert und im neu anzulegenden Verzeichnis „<Stamm>/01“ fortgesetzt. Um eine entsprechende Skalierung zu gewähr-

leisten, ist es also angebracht, von vornherein einen ausreichend großen Adressraum zur Verfügung zu stellen.

Die Aufteilung des Pfades in jeweils zwei Stellen erfolgt aus dem Grund, dass viele Dateisysteme nicht in der Lage sind, größere Mengen an Dateien in einem Verzeichnis in logarithmischer Zeit zu verwalten. Ein einfaches „ls“ in einem ext2-Dateisystem und 100 000 Dateien in einem Verzeichnis führt zu einer beachtlichen Wartezeit. Mit Blick auf die geforderte Portierbarkeit muss von anderen, noch deutlich weniger leistungsfähigen Dateisystemen wie FAT32 ausgegangen werden. So befinden sich in jedem Verzeichnis lediglich maximal 256 Unterverzeichnisse – eine Zahl, die jedes Dateisystem mit ausreichender Performanz verwalten kann.

Die Ordnung nach Pfad-Wort ist aus dem Grund so gewählt, weil es XML-Tags gibt, die keinen Content enthalten. Sie haben keinen schließenden Tag und stehen für sich. In diesem Fall ist der Pfad zur „LL“-Datei: “<Stamm>/3f/4d/00/00/00/LL“. Mit einer Anordnung, in welcher der Pfad zu einer Pfadnummer, der mangelnde Content aber nur zu einer „0“ führt, sind derartige Pfadelemente ohne Probleme abbildbar.

Die eingangs geforderte Eigenschaft, dynamische ‚Drilldowns‘ anzubieten, wird mit folgendem Mechanismus unterstützt: Ein Drilldown besagt, dass in der Ergebnismenge eines Suchworts alle erkannten Entitäten dynamisch und in Abhängigkeit vom Suchwort ausgegeben werden sollen. Wenn beispielsweise nach „HSV“²⁴ gesucht wird, sollen alle Namen der Fußballspieler ausgegeben werden, die ebenfalls, und zwar ausschließlich, in den Dokumenten vorkommen, in welchen der gesuchte String (hier „HSV“) vorkommt. In Bezug auf semistrukturierte Dokumente bedeutet das, dass der Spielernamen in der Dokumentenverarbeitung gesondert als solcher ausgezeichnet wird und die Wortindex-Nummer aller Spielernamen in demselben Verzeichnis, in welchem sich die „LL“-Datei befindet, in einer „ZZ“-Datei verzeichnet wird. Entsprechend der Häufigkeit kann hier bereits eine erste Sortierung vorgenommen werden.

Übertragen auf die historische Fragestellung und unter der Voraussetzung, dass eine Eigennamenerkennung in der Dokumentenverarbeitung zuverlässige Ergebnisse liefert, kann damit etwa die Frage beantwortet werden, welche Personen in einem bestimmten historischen Kontext auftauchen. Falls man eine Ortsnamenerkennung implementiert, kann in Bezug auf einen bestimmten historischen Vorgang ein Profil aller Orte erstellt

²⁴ Hamburger Sportverein.

werden, zu denen der Vorgang in Bezug steht. Ein Drilldown würde nun in der ersten Ergebnisdarstellung die erkannten Entitäten in Bezug auf ein Suchwort ausgeben.

Der Benutzer würde in einem nächsten Schritt den ausgewählten Begriff zusammen mit dem ursprünglichen Suchwort als erneute Suche abschicken. Ab hier handelt es sich um Suchvorgänge mit mehreren Constraints, sodass Joins unvermeidlich sind. Die Bearbeitung von Joins von Ergebnismengen ist jedoch ein weiterführendes Thema und soll an dieser Stelle explizit als offener Punkt vermerkt werden. In dieselbe Problematik fällt auch die der Verarbeitung von Seitenbedingungen, die in XPath spezifiziert sind.

Wichtig in Bezug auf die Erfassung von Drilldowns im Index ist, dass die Auszeichnung einer Entität außerhalb des Index erfolgt. Der Index selbst kann ohne eine erforderliche Neukonfiguration und ein erneutes Einlesen des gesamten Corpus ohne Unterbrechung weiter betrieben werden. Die Auszeichnung erfolgt in einem Modul in der Dokumentenverarbeitung und kann beliebig vom Benutzer konfiguriert werden.

4.7 Der QueryProcessor

Der QueryProcessor übernimmt vom QueryHandler, der Steuerkomponente der Query-Verarbeitung, eine Strukturinformation und ein Eingabewort in Form eines standardkonformen XPath-Ausdrucks²⁵ sowie einen auf dem Eingabewort fertig kompilierten Kontrollautomaten, der implizit die maximal erlaubte Approximationsdistanz bereits enthält. Seine Aufgabe ist, die Anfrage auf der Basis der zuvor vorgestellten Datenstrukturen abzuarbeiten und eine Liste an treffenden Locations zurückzugeben, wobei jede Location mit dem Multiplikator der approximativen Distanz, aber auch den TF/ICF-Werten annotiert wird. Die Auswertung der für das Ranking relevanten Informationen erfolgt durch den Ranker.

Nach der Übernahme des XPath-Ausdrucks wird dieser in seinen Content- und seinen Strukturanteil aufgetrennt. Zudem wird ermittelt, ob der Strukturanteil Freiheitsgrade enthält oder nicht.

Der mit übergebene Automat dient als Kontrollsystem bei der Bearbeitung des WordIndex, wobei in beiden Strukturen parallel in der Tiefe fortgeschritten wird. Falls ein Zeichen aus dem Eingabewort nicht weiter im

²⁵ Der XPath-Standard. Adresse: <http://www.w3c.org/TR/xpath> (letzte Einsichtnahme am 17.05.2006).

WordIndex vorhanden ist, wird anhand des Kontrollautomaten festgestellt, ob die bisher vorgenommenen Abweichungen noch innerhalb der maximal zulässigen Editierdistanz liegen. Wenn dies der Fall ist, kann im WordIndex nach einem alternativen Übergang gesucht und dieser Weg weiter verfolgt werden. Ein Treffer ist dann gefunden, wenn sowohl im Kontrollautomaten als auch im WordIndex ein Endzustand erreicht wird. Allerdings endet die Suche in diesem Fall nicht nach dem Finden eines ersten Treffers, sondern mittels Backtracking wird versucht, den gesamten Suchraum innerhalb der definierten Fehlertoleranz auszuschöpfen.

Da ein Treffer im WordIndex zu einem Lametta-Bitstring führt, ist nun auch eine Abarbeitung der Überprüfung des Strukturanteils möglich. Falls der Strukturanteil keine Freiheitsgrade aufweist, so kann direkt mit einem Containment-Test ermittelt werden, ob die Anfrage erfüllbar ist. Falls Freiheitsgrade vorhanden sind, so wird anhand des Government-Tests im Lametta-Bitstring untersucht, wie weit in den Dataguide-Baum hinabgestiegen werden kann. Ein positiver Government-Test besagt, dass Kinder des aktuell betrachteten Knotens im Corpus das besagte Suchwort enthalten. Bei jedem Kind wird dann wiederum ein Containment-Test vorgenommen, um Knoten zu identifizieren, in denen im Corpus das zuvor identifizierte Wort auch wirklich vorkommt.

Führen Government- sowie Containment-Tests anhand des Lametta-Bitstrings schließlich zu Knoten, mit denen die Anfrage erfüllbar ist, so wird aus den von den beiden Datenstrukturen erhaltenen Nummern der Pfad zur Locations-Datei zusammengestellt. Die Locations werden zusammen mit der erreichten Editierdistanz und mit den einzeln verfügbaren TF/ICF-Werten an den Ranker gegeben.

4.8 Die Query Processing Pipeline

Grundsätzlich ist jede Indexierung ein symmetrisches Verfahren: Damit die Dokumente in einem Corpus gefunden werden können, müssen die Texte in die Einheiten zerlegt werden, die später gefunden werden wollen. Also enthält ein Index atomare Informationsentitäten, die aus den zu indexierenden Dateien oder Dokumenten herausgearbeitet werden müssen. Die Symmetrie besteht insofern, als auch die Anfrage an eine Suchmaschine einer Verarbeitung unterworfen sein muss: Die Anfrage muss so weit aufbereitet werden, dass sie den indexierten Entitäten exakt entspricht. Archi-

tektonisch ergeben sich daher zwei Verarbeitungspipelines, eine so genannte ‚Document Processing Pipeline‘ sowie eine ‚Query Processing Pipeline‘.

Der Unterschied zwischen beiden Zugriffsmethoden auf den Index besteht darin, dass die Query Processing Pipeline die Verarbeitung in Echtzeit vornehmen muss, und die gefundenen Ergebnisse noch einer Nachbearbeitung unterworfen werden müssen. In letzterer muss noch eine Gewichtung und die Auswertung der Rankingparameter errechnet werden. Weil die Document Processing Pipeline keinen zeitkritischen Einschränkungen unterworfen ist, da die Dokumentenverarbeitung vorab erfolgt, hängt eine gute Antwortzeit direkt von der Verarbeitungsgeschwindigkeit der Query Processing Pipeline ab. Im Gegensatz zur Dokumentenverarbeitung findet man in Installationen kommerzieller Suchmaschinen bei der Anfrageverarbeitung selten Module in Skriptsprachen. Allerdings sind die Aufgaben auch meist deutlich weniger komplex.

Die Query-Verarbeitung hat die Aufgabe, aus dem eingegebenen String und der in der GUI ausgewählten Strukturinformation alle nötigen Informationen zu extrahieren und diese dergestalt aufzubereiten, dass sie an den Index übergeben werden können. Komponenten einer typischen Query Processing Pipeline sind:

- Encoding-Feststellung,
- Normalisierung auf UTF-8,
- Erkennung der eingegebenen Sprache,
- Tokenisierung,
- Lemmatisierung.

Da die Entwicklung derartiger Module einen nicht zu unterschätzenden Mehraufwand darstellt, soll im Rahmen dieses Projekts lediglich ein Tokenisierungsmodul entwickelt werden. In Bezug auf die übrigen Arbeitsschritte ist die Entwicklung der entsprechenden Module an den Bedarf aus dem zukünftigen Einsatz gebunden. Da während der Entwicklungsphase des vorliegenden Projekts von einer kontrollierbaren Datenlandschaft und sauberen Beispieldaten ausgegangen werden kann und ein entsprechender Bearbeitungsbedarf nicht besteht, werden die Module von passiven Prototypen ersetzt, die später gegen funktionale Komponenten ausgetauscht werden können.

Das Ergebnis der Anfragebearbeitung ist in jedem Fall ein standardkonformer XPath-Ausdruck sowie ein kompilierter Kontrollautomat, der die Information enthält, ob und in welchem Rahmen eine approximative Suche vorgenommen werden muss. Sollte der Eingabestring der Anfrage aus mehreren Wörtern bestehen, so werden im Query-Tokenizer mehrere XPath-Anfragen mit gegebenenfalls mehreren Automaten generiert. Die Schnittstelle zum Index ist explizit konform zum XPath-Standard, um auch maschinelle Anfragen aus anderer Quelle abarbeiten zu können.

4.9 Der Kontrollautomaten-Compiler

Je nach Anforderung und Approximationsbedarf wird nach dem Verfahren von Mihov und Schulz auf jedes Eingabewort ein Automat kompiliert, der gemäß der Definition der Levenshtein-Editieroperationen²⁶ (Einfügungen, Löschungen, Ersetzungen) die Kontrolle von Abweichungen innerhalb bestimmter Distanzgrenzen erlaubt.²⁷

Der Kompilierungsvorgang verläuft jedoch nicht rein algorithmisch, sondern es werden äquivalente Übergänge einkompiliert, sodass etwa Phänomene wie die Äquivalenz von „u“ und „v“ im Anlaut ohne Feststellung eines Fehlers behandelt werden können. Es können so aus der diachronen Linguistik bekannte graphemische (Allographe) oder phonetische Phänomene abgebildet und entsprechend behandelt werden.²⁸ Ziel hierbei ist die mit Blick auf die Eigenschaften der alten Texte erforderliche Erweiterung des Recall.

4.10 Der Ranker

Wie oben beschrieben, trennen wir zwischen der Bewertung der Contentrelevanz und der Bewertung der Strukturerelevanz: Beinhaltet die Strukturinformation keine Freiheitsgrade, so ist eine Gewichtung der Strukturinfor-

²⁶ Vladimir I. Levenshtein: Binary codes capable of correcting deletions, insertions, and reversals. In: Doklady Akademii Nauk SSSR 163 (1965). S. 845–848 (Russisch). Englische Übersetzung in: Soviet Physics Doklady 10 (1966). S. 707–710.

²⁷ Vgl. Kemal Oflazer: Error-tolerant finite-state recognition with applications to morphological analysis and spelling correction. In: Computational Linguistics 22/1 (1996). S. 73–89.

²⁸ Zu den diachronen Varianzen zum und vom Frühneuhochdeutschen in spätere Sprachstände: Gerhard Philipp: Einführung ins Frühneuhochdeutsche. Sprachgeschichte, Grammatik, Texte. Heidelberg 1980.

mation überflüssig, da der Benutzer dem System keinen Entscheidungsspielraum lässt. Beinhaltet die Strukturinformation doch Freiheitsgrade, so erfolgt in unserer Architektur eine benutzerbezogene Relevanzgewichtung: Unter der Annahme, dass es sich um eine zweckgebundene Suchmaschine für eine bestimmte Benutzergruppe und für eine bestimmte Sorte an Dokumenten handelt, sehen wir die Möglichkeit der Gewichtsmultiplikation durch den Ranker vor. Hier genügen wir der Forderung, dass verschiedene Nutzergruppen einer Suchmaschine für historische Daten unterschiedliche Interessenschwerpunkte haben können. Indem abhängig vom Benutzerprofil eine besondere Ergebnismenge vorgenommen werden kann, lässt sich die Ergebnismenge so sortieren, dass beispielsweise eher Treffer aus dem Originaltext oder eher Treffer aus dem Kommentar oder eher Treffer aus den Metadaten angezeigt werden.

In die Errechnung der Sortierung der Locations fließen damit die Editierdistanz aus der Abarbeitung des Automaten, der TF/ICF-Wert sowie die benutzerorientierte Gewichtung der Strukturinformation ein.

5. Projektstand

Zum Zeitpunkt der Kalenderwoche 16/2006 sind der Feeder samt XML-Parser, die Architektur der DocumentProcessing Pipeline, der DataGuide und erste Teile des WordIndex fertig gestellt.

Digitale Erschließung und Sicherung von aktuellen archäologischen Befunden

Christoph Schäfer

1. Forschungsdiskussion im Niemandsland

Vehement vorgebracht und erbittert ausgetragen haben erst jüngst wieder einige Forschungskontroversen aus den Altertumswissenschaften nicht allein die Fachwelt, sondern gerade auch die Öffentlichkeit erschüttert. Als Beispiele sei nur der ‚neue Streit um Troja‘ angeführt, in dem die Protagonisten, bezeichnenderweise beide Professoren an der Universität Tübingen, sich alsbald auf einer wenig sachlichen Ebene äußerten.¹

Kaum weniger aufgeladen ist die Diskussion um die Varusschlacht. Insbesondere bezüglich der Befunde des frühkaiserzeitlichen Schlachtfelds von Kalkriese schlagen die Wogen hoch. Interviews und Artikel etwa im Spiegel sorgen auch hier für ein ausgesprochen intensives Interesse der Öffentlichkeit. Will man sich ein wirklich fundiertes Bild von den aktuellen Ergebnissen machen, ist man allerdings auf die Publikation von Grabungsberichten angewiesen, die in der Regel erst Jahre später, oftmals auch gar nicht publiziert werden. Das Fehlen aktuellster Informationen erschwert und verzögert nicht allein eine solide Meinungsbildung beim Fachwissenschaftler, es macht es auch nahezu unmöglich, auf solche aktuellen Probleme und Tendenzen in der akademischen Lehre zu reagieren. Eine adäquate Stellungnahme oder ein Reagieren auf Anfragen aus Kreisen von Nichtspezialisten ist so kaum möglich.²

¹ Vgl. Christoph Ulf (Hg.): Der neue Streit um Troja. Eine Bilanz. München 2004.

² Rainer Wiegels und Winfried Woesler (Hg.): Arminius und die Varusschlacht. Geschichte – Mythos – Literatur. Paderborn u.a. 2003. – Rainer Wiegels (Hg.): Die Fundmünzen von Kalkriese und die frühkaiserzeitliche Münzprägung. Akten des wissenschaftlichen Symposions in

2. Bilder in der (Alttertums-)Wissenschaft

Genau dieses Problem steht im Mittelpunkt eines Projekts an der Universität Hamburg, das die möglichst zeitnahe multimediale Dokumentation von zentralen archäologischen Befunden und deren alttertumswissenschaftliche Einbettung zum Ziel hat. Dabei geht es zum einen gerade darum, den geschilderten Defiziten entgegenzuwirken und über Interviews sowie die Aufnahme von Befunden einen deutlich *schnellere*n Austausch von neuen Erkenntnissen zu ermöglichen. Darüber hinaus lässt sich mit entsprechender Software aufgrund der erfassten Daten ein Zwischenstand visualisieren, der wiederum zu weiteren Forschungsdiskussionen anregen kann. Speziell für den Einsatz in der Lehre gilt es, die suggestive Kraft der Bilder durch entsprechende Kommentare und spezifische Quellenkritik zu flankieren. Außerdem müssen je nach Zielpublikum zu jedem Themenbereich einzelne Sequenzen didaktisch aufbereitet werden. Der Nutzer kann nicht allein gelassen werden mit dem Rohmaterial.³

Bilder haben schon immer ihren Sitz in der Wissenschaft. Sie dienen als Quellen wissenschaftlicher Erkenntnis, als methodische Hilfsmittel, zum Sichtbarmachen des Nichtsichtbaren und nicht zuletzt als Instrument zur Wissensvermittlung.⁴ „Durch wissenschaftliche Zeichnungen, Graphen, Diagramme, Computerprogramme, Photographien, Röntgenbilder, Kartographien etc. wird wissenschaftliches Wissen überhaupt erst in eine stabile, aussagekräftige und kommunizierbare Form gebracht.“⁵ Eben diese Eigenschaft zur Kommunikation ist auch Voraussetzung für den Wissenstransfer. Längst sind Bilder, die über den rein wissenschaftlichen Bereich hinausge-

Kalkriese, 15.–16. April 1999 (Osnabrücker Forschungen zu Altertum und Antikerezeption 3). Mönese 2000.

³ Didaktischer Mehrwert entsteht nicht einfach durch Medieneinsatz allein, sondern nur durch sorgfältiges didaktisches Design. Vgl. Günter Dörr und Peter Strittmatter: Multimedia aus pädagogischer Sicht. In: Ludwig J. Issing und Paul Klimsa (Hg.): Information und Lernen mit Multimedia und Internet. Weinheim 2002. S. 144 – Carmen Ueckert: Interaktionsstrukturen universitärer Lehre mit neuen Medien. In: Bernd Pape, Detlev Krause und Horst Oberquelle (Hg.): Wissensprojekte. Gemeinschaftliches Lernen aus didaktischer, softwaretechnischer und organisatorischer Sicht (Medien in der Wissenschaft 27). Münster u. a. 2004. S. 128–147.

⁴ Dazu jetzt vor allem Gyula Pápay: Bilder der Wissenschaft – Wissenschaft der Bilder. Einleitende Bemerkungen zur allgemeinen Bildwissenschaft. In: Ders. (Hg.): Bilder der Wissenschaft – Wissenschaft der Bilder. Rostock 2005. S. 7–18.

⁵ Monika Dommann und Mariette Meiner: Wissenschaft, die Bilder schafft. In: Dies. (Hg.): Wissenschaft, die Bilder schafft. Zürich 1999. S. 15.

hen, zum Untersuchungsgegenstand geworden, eine allgemeine Bildwissenschaft ist in der Genese begriffen.⁶

Der Siegeszug des Bildes ist kaum aufzuhalten. Während bei Platon Bilder dem abstrakt-begrifflichem Denken noch klar untergeordnet werden, gewinnt heute das Bild dem schriftlich geäußerten Gedanken gegenüber oft genug die Oberhand. Man mag dies beklagen, die Überflutung durch die Medien geißeln und wie Jean Baudrillard Bilder als „Mörder des Realen“ bezeichnen, dennoch kommt man nicht umhin, sich der Herausforderung durch die Dominanz des Visuellen zu stellen.⁷ In ihrer einschlägigen Untersuchung kommt Barbara M. Stafford zu dem Ergebnis, dass in unserer Kultur das Bild die Schrift bereits an Bedeutung überflügelt hat.⁸ Seit den 1990er Jahren charakterisiert man diese Entwicklung regelrecht als ‚imagic turn‘, ‚pictorial turn‘ oder ‚iconic turn‘.⁹

Wie ist hier die Altertumswissenschaft aufgestellt? Kurz gesagt: schlecht! Abgesehen von der Nutzung digitaler Techniken für den ganz in konservativen Bahnen verlaufenden Umgang mit Bildmaterial (Stichwort ‚Digitale Diathek‘ oder ‚Bilddatenbank‘ zum Beispiel archäologischer Objekte) tut sich nur allmählich etwas in Richtung eines bewussten Bemühens um die Gestaltung von Bildern. Meist werden wirkmächtige Animationen von Medienspezialisten, und zwar nicht zuletzt den Herstellern von Computerspielen, entwickelt, die über keinerlei fachlichen Hintergrund verfügen.

Hier gilt es dringend, den Schritt aus dem Elfenbeinturm zu wagen, sonst verlieren wir auch den Rest an Einfluss auf die Bilder, die den Wissenstransfer in breiten Kreisen der Bevölkerung schon heute dominieren. An diesem Punkt setzt das Projekt *Römer und Germanen – Konfrontation und Integration* an.

Unter wissenschaftlichen Gesichtspunkten sollten die Chancen filmischer Dokumentation sowie der Entwicklung von 3D-Rekonstruktionen und Animationen aller Art genutzt werden, um eine Beschleunigung und Dynamisierung der Fachdiskussion zu ermöglichen. Über die Medientechnologien

⁶ Klaus Sachs-Hombach: Die Magdeburger Bemühungen um eine allgemeine Bildwissenschaft. In: Gyula Pápay (Hg.): *Bilder der Wissenschaft – Wissenschaft der Bilder*. Rostock 2005. S. 41–51.

⁷ Vgl. Hans Belting: *Bild-Anthropologie*. München 2001. S. 18.

⁸ Barbara M. Stafford: *Kunstvolle Wissenschaft, Aufklärung, Unterhaltung und der Niedergang der visuellen Bildung*. Amsterdam, Dresden 1998.

⁹ Pápay, wie Anm. 4, S. 10.

nik lassen sich etwa archäologische Befunde auch im Zwischenstand dokumentieren, man muss nicht viele Jahre oder Jahrzehnte auf Grabungspublikationen warten, sondern kann sofort buchstäblich ‚in medias res‘ gehen. Ein zusätzlicher Effekt ergibt sich aus der schlaglichtartigen Aufnahme, die unter Umständen wissenschaftshistorisch höchst interessante Einblicke in die Genese der Erkenntnisse oder einer Lehrmeinung zu einem umstrittenen Sachverhalt bieten kann. So kann das Material selbst zur Quelle für wissenschaftsimmanente Prozesse werden.



Abbildung 1: Interview mit Dr. Andreas Boos (Regensburg) zu aktuellen Grabungsergebnissen.

Prinzipiell ändert auch der kreative Umgang mit digitalen Techniken und Speichermedien nichts an den Methoden des Historikers. Mediendaten müssen selbstverständlich mit den üblichen historischen Quellen kombiniert werden. Heuristik und Hermeneutik büßen nichts von ihrer Bedeutung ein. Bei der Darstellung der Untersuchung gilt es jedoch noch mehr Aspekte zu berücksichtigen als bisher. Unabdingbar ist die intensive Auseinandersetzung mit gestalterischen und didaktischen Prinzipien, keinesfalls kann man Rohmaterial unbearbeitet aus der Hand geben, ansonsten

wirkt die lieblose Bearbeitung wie ein negativer Kommentar, selbst wenn man dies gar nicht beabsichtigt.

Dringend notwendig wäre eine Diskussion über Archivierungs- und Nutzungskonzepte solcher Daten. Videoserver können eine wichtige Basis bieten. Derzeit erlaubt der Datendurchsatz im WWW noch keine adäquate Nutzung von Videodaten.¹⁰ In dem Hamburger Projekt fiel daher die Entscheidung für die DVD als Medium zur Verbreitung und dies lag nicht zuletzt an den weiteren Facetten des Projektes, das neben der wissenschaftlichen Seite im Sinne der Nachhaltigkeit auf ein breites Publikum und eine (teilweise) Refinanzierung durch Vermarktung zielte.

3. „Römer und Germanen – Konfrontation und Integration“

Schon bei den Vorüberlegungen zum Projekt *Römer und Germanen* wurde auf zeitgemäße Bezüge, den ‚Sitz im Leben‘, geachtet. Erleichtert wurde dieses Bemühen, weil aktuelle Forschungsergebnisse zum Thema ‚Konfliktlösung und Integration im Zusammenleben von Römern und Germanen‘ eine teils frappierende Ähnlichkeit antiker und heutiger Probleme im Umgang mit Fremden erkennen lassen. Im Rahmen der inzwischen fertiggestellten Multimedia-Anwendung werden Problemlösestrategien aufgezeigt und Denkanstöße für die heutige Zeit gegeben. Der kombinierte Einsatz neuer Techniken mit einem handlungsorientierten Lernkonzept trägt dazu bei, dass auch die vermittelten althistorischen Inhalte und die Komplexität historischer Prozesse von den Nutzern dauerhaft memoriert werden.

Im Einzelnen geht es um den effizienten Wissenstransfer mittels interaktiver Multimedia-Anwendungen, die ein zeitgemäßes Erlernen der fachlichen Inhalte ermöglicht. Die Visualisierung erfolgte vor allem über Filmsequenzen, die nicht allein dokumentarische Zeugnisse, sondern zur besseren Veranschaulichung auch Szenen an Originalschauplätzen enthalten. Ziel war es, einerseits ein Lern- und Trainingsprodukt für den Erwerb zen-

¹⁰ In anderen Bereichen der Wissenschaft wird – wenn auch ohne die spezielle Ausrichtung auf Videodaten – die Entwicklung von ähnlichen Konzepten und deren Umsetzung in die Praxis energisch vorangetrieben. Eine solche Entwicklung wurde etwa am Institut für Multimediale und Interaktive Systeme (IMIS) in Lübeck realisiert. Vgl. Barbara Kleinen: Wissensarchive in der Praxis. In: Bernd Pape, Detlev Krause und Horst Oberquelle (Hg.): Wissensprojekte. Gemeinschaftliches Lernen aus didaktischer, softwaretechnischer und organisatorischer Sicht (Medien in der Wissenschaft 27). Münster u. a. 2004. S. 158–174.

traler Kenntnisse über die innere Struktur und Gesellschaft des Imperium Romanum zu erstellen, andererseits aber auch das Thema ‚Fremdenfeindlichkeit und Integration von Minderheiten‘ exemplarisch in neuem Licht zu behandeln.

Insbesondere durch die Dreharbeiten an gegenwärtigen Brennpunkten der Archäologie und Alten Geschichte konnten aktuellste Forschungsergebnisse gesichert und aufbereitet werden, deren anderweitige Publikation noch auf Jahre hinaus auf sich warten lassen wird.

Schon bei den Planungen spielte der Wissenstransfer über die Grenzen der eigenen Universität und der Wissenschaft an sich hinaus eine wesentliche Rolle. Daher wurde schon im Vorfeld die Vermarktung der Ergebnisse ins Auge gefasst und eine diesbezügliche Kooperationsvereinbarung mit einem kompetenten Partner aus der Wirtschaft ausgehandelt. Im Einsatz digitaler Techniken konnten neue Felder erschlossen werden: 3D-Animationen wurden angefertigt und für die Nachbearbeitung des Tons die School of Audio Engineering (SAE) in Hamburg gewonnen. In den Bereichen Entwicklung und Vertrieb ist eine enge Kooperation mit dem Institut für Film und Bild in Wissenschaft und Unterricht in München (FWU) Kernstück des Projekts.

Aus den erarbeiteten Materialien wurden Lerneinheiten für die inneruniversitären Lernplattformen geschaffen sowie eine didaktische DVD mit DVD-ROM-Teil für den Einsatz in der Schule und ein breiteres Publikum erstellt. Marketingstrategien sind vorhanden und wurden mit dem FWU sowie dem Multimedia-Kontor Hamburg (MMKH) als weiterem Partner abgestimmt.

Inzwischen wurde die DVD erfolgreich im Beisein von Journalisten regionaler wie überregionaler Tageszeitungen sowie von Radio- und Fernsehreportern in der Gelehrtenschule des Johanneums in Hamburg sowie in weiteren Schulen in Bayern und Rheinland-Pfalz getestet. Das überaus positive Echo in der Presse ließ die Absatzzahlen hochschnellen.



Abbildung 2: Cover der DVD „Römer und Germanen – Konfrontation und Integration“ (2005).

4. Perspektiven

Was die Perspektiven angeht, so eröffnet die professionelle Produktion einer didaktischen DVD in Kooperation mit einem Partner aus der Wirtschaft sowohl im Hinblick auf die Kompetenz des Teams als auch auf den Bild und Filmstock enormes Potential für weitere Produktionen, die angesichts der Vorarbeiten deutlich preisgünstiger realisiert werden können. Mittelfristig kann sogar eine Kostendeckung angestrebt werden.

Dabei ergibt sich sowohl in der Forschung als auch in der Lehre (Einsatz des Materials auf der Lernplattform, zusätzliche Lehrveranstaltungen zu fachspezifischer EDV) ein effektiver Zugewinn an Ressourcen.

Digitale Urkundenbücher zur mittelalterlichen Geschichte

Jürgen Sarnowsky

Die digitalen Medien haben international für die wissenschaftliche Arbeit in den Geisteswissenschaften mittlerweile grundlegende Bedeutung erlangt, so dass in diesem Kontext dafür immer häufiger zu Recht von E-Science – oder mit einem Begriff von Manfred Thaller: ‚E-Humanities‘ – die Rede ist, auch in Abgrenzung vom E-Learning. Der virtuelle Besuch in Bibliotheken mit Katalogen, Datenbanken und digitalen ‚Nachschlagewerken‘ ist inzwischen ebenso selbstverständlich wie die Heranziehung digitalen Materials in multimedialer Form, sei es über das World Wide Web, sei es auf CD oder DVD. Dazu gehört auch ein wachsendes Angebot an digitalen Urkundenbüchern, nicht zuletzt für den Bereich der mittelalterlichen Geschichte.¹ Die Bearbeitung des Materials geschieht auf zweifache Weise: Zumeist werden die gedruckten Urkundenbuch-Editionen – in unterschiedlicher Form – digitalisiert, erschlossen und durch weiteres Material ergänzt, teilweise erfolgt aber auch eine Neuaufnahme von Quellen. Schon angesichts der Tatsache, dass herkömmliche Urkundenbuchprojekte bei breiterer Überlieferung aus dem späteren Mittelalter vielfach ins Stocken geraten sind, bieten digitale Publikationsformen eine denkbare Alternative. Technisch finden sich inzwischen die verschiedensten Lösungen: die bildliche Wiedergabe bereits gedruckter Sammlungen, teilweise begleitet von digitalen Textfassungen; einfache HTML- oder auch XML-Lösungen mit hierarchischen Strukturen und sukzessiver Erschließung von Materialien; immer wieder erweiterte Datenbank-Lösungen, und vieles andere mehr. Die Quellen sind teilweise durch Abbildun-

¹ Eine gute Übersicht bietet die von Georg Vogeler und anderen betreute Seite zu den Historischen Hilfswissenschaften, Urkundenbüchern und Regestensammlungen. Adresse: <http://www.vl-ghw.uni-muenchen.de/chartularia.html> (letzte Einsichtnahme am 25.03.2006).

gen von Originalen und Karten ergänzt; sie werden durch Suchmaschinen erschlossen.

Seit einigen Jahren wird vor diesem Hintergrund eine intensive Diskussion um Standards geführt, nicht zuletzt um eine angemessene Auszeichnung bzw. Indexierung der Dokumente,² ebenso aber um die sinnvolle Präsentation des Materials in digitaler Form.³ Insbesondere die bildliche Wiedergabe der alten Druckausgaben hat die Anlehnung an das Erscheinungsbild von Drucken gefördert, so dass eine Reflexion über die Möglichkeiten von digitalen Urkundenbüchern und die Anforderungen an sie notwendig ist. Dies soll in diesem Beitrag gewissermaßen aus doppelter Perspektive geschehen: aus der des Bearbeiters von drei digitalen Urkundenbüchern, des *Virtuellen Preußischen Urkundenbuchs*, des *Virtuellen Hamburgischen Urkundenbuchs* und der international angelegten *Hospitaller Sources*⁴ sowie aus der des Benutzers. Ich werde dabei an Überlegungen anschließen, die ich vor einem Jahr auf einer Tagung in Mainz vorgestellt habe.⁵ Technische Aspekte sollen hier eher zurückstehen, ebenso die

² Zur Münchner Tagung von 2004 und der daraus hervorgegangenen Arbeitsgruppe vgl. die Übersicht unter Adresse: http://www.geschichte.uni-muenchen.de/ghw/UrkdTD_Expose.shtml (letzte Einsichtnahme am 27.03.2006), die Startseite der Arbeitsgruppe (*Charters Encoding Initiative*) unter Adresse: <http://www.cei.lmu.de> (letzte Einsichtnahme am 27.03.2006) mit dem Vorschlag für eine ‚tag library‘ sowie Georg Vogeler: Ein Standard für die Digitalisierung mittelalterlicher Urkunden mit XML. Bericht von einem internationalen Workshop in München 5.–6. April 2004. In: Archiv für Diplomatik 50 (2004). S. 23–34. – Vgl. auch die Beiträge von Markus Heller und Georg Vogeler in diesem Band.

³ Dazu siehe insbesondere Patrick Sahle und Georg Vogeler: Urkundenforschung und Urkundenedition im digitalen Zeitalter. In: Geschichte und Neue Medien in Forschung, Archiven, Bibliotheken und Museen. Tagungsband *.hist2003*. Hg. von Daniel Burkhardt, Rüdiger Hohls und Vera Ziegeldorf (Historisches Forum. Schriftenreihe von Clio-online 7,1). Berlin 2005. S. 333–382; vgl. die Online-Fassung: http://edoc.hu-berlin.de/e_histfor/7_1/PHP/Publikationen_7-2005-I.php#005008 (letzte Einsichtnahme am 27.03.2006).

⁴ Zum Stand vgl. Jürgen Sarnowsky: Hospitaller Sources – a Project for a Source Book in the Internet. In: Bulletin of International Medieval Research 8 (2002, erschienen 2003). S. 13–20. – Ders.: Das virtuelle Hamburgische Urkundenbuch – ein digitales Editionsprojekt. In: Hansische Geschichtsblätter 121 (2003). S. 161–170. – Ders.: Das virtuelle Preußische Urkundenbuch – neue Wege der Kooperation für Internet-Editionen, revidierte Fassung. In: Beiträge zur Geschichte Westpreußens 19 (2004, erschienen 2005). S. 257–266.

⁵ Auf dem Workshop *Buch und Internet – Aufbereitung historischer Quellen im digitalen Zeitalter* der Regesta Imperii an der Akademie der Wissenschaften und der Literatur in Mainz, abstract(s) unter Adresse: <http://regesta-imperii.uni-giessen.de/downloads/onlinedokumente/workshopmaterialien.pdf> (letzte Einsichtnahme am 27.03.2006).

sich im Bereich E-Humanities bildenden Strukturen.⁶ Ausgehend von einigen allgemeineren Überlegungen in neueren Publikationen zum Thema, sollen in diesem Beitrag zunächst einige Urkundenbücher im Internet vorgestellt und analysiert werden. Daran sollen sich eine Auswertung und die Formulierung von Forderungen und Möglichkeiten für Weiterentwicklungen anschließen.

Am Beispiel des *Hansischen Urkundenbuchs* haben Patrick Sahle und Thorsten Schaßan schon vor einigen Jahren, in einem im Jahr 2000 erschienenen Aufsatz, ihre Vorstellungen und Anforderungen an ein Urkundenbuch im Internet formuliert.⁷ Dabei haben sie eingangs folgende Erwartungen formuliert:

- Die digitale Version des Urkundenbuchs soll die Quellen besser nutzbar und verfügbar machen;
- sie soll „das in den Texten, Regesten, Anmerkungen oder Registern, teilweise aber auch in den Layoutstrukturen vorhandene Wissen in einer Weise explizit werden lassen, die neue Zugriffswege und Nutzungsformen erlaubt“;⁸
- sie soll offen sein für jede Art von Ergänzungen und Änderungen;
- sie soll so zugleich dazu anregen, dass die Strukturen und Funktionen derartiger Quellensammlungen überprüft und gegebenenfalls modifiziert werden.

Um diese Ziele zu erreichen, soll ein mehrstufiger Prozess durchlaufen werden, der von der automatisierten Textaufnahme durch Texterkennung über die Ergänzung zusätzlicher struktureller und inhaltlicher Informationen zu einem ‚offenen Informationssystem‘ führen soll. Bei der Texterfassung ist auf die Übertragung des Layouts zu achten, da die Formatie-

⁶ Vgl. zum Beispiel das Trierer ‚Kompetenzzentrum für elektronische Erschließungs- und Publikationsverfahren in den Geisteswissenschaften‘, <http://germazope.uni-trier.de/Projects/KoZe2> (letzte Einsichtnahme am 27.03.2006). In Hamburg das Zentrum ‚Geisteswissenschaften in der digitalen Welt‘, <http://www.phil-gesch.uni-hamburg.de/gdig/gdigstart.html>, und die *Hamburg Digital Humanities*, <http://www1.uni-hamburg.de/narratologie/hdh/index.html>.

⁷ Patrick Sahle und Thorsten Schaßan: Das Hansische Urkundenbuch in der digitalen Welt. In: *Hansische Geschichtsblätter* 118 (2000). S. 133–155.

⁸ Sahle/Schaßan, wie Anm. 7, S. 133. – Vgl. auch die allgemeinen Überlegungen in Patrick Sahle: Vom editorischen Fachwissen zur digitalen Edition. Der Editionsprozeß zwischen Quellenbeschreibung und Benutzeroberfläche. In: *Fundus. Forum für Geschichte und ihre Quellen* 2 (2000). S. 75–102.

rungen des Drucks immer schon eine Bedeutung transportieren. Texte und Textformate sollen in eine einfache, maschinenlesbare Form gebracht und so vielfältig, bis hin zur Generierung von Wörterbüchern, nutzbar gemacht werden. Ergänzungen wären dann durch interaktive Verschlagwortung durch die Nutzer, Vernetzung mit weiteren Informationsquellen, Korrekturen von Fehlern, Aufnahme anderer Fassungen und Abbildungen der Quellen, Integration von Sekundärliteratur und Anderes zu erreichen. Damit wäre die – ohnehin nicht stringente – Ordnung des gedruckten Urkundenbuchs durchbrochen, vielmehr würde sich die Frage der aufzunehmenden Texte (was sind ‚Urkunden‘) und ihrer Strukturierung neu stellen. Auch die Bearbeiter (und der Bearbeitungsstand) wären zu berücksichtigen. Wenn dabei Schwächen der gedruckten Fassung – die unterschiedlichen Standards und Auswahlkriterien der Bearbeiter und Anderes mehr – ausgeglichen werden sollen, bedarf es aber der ‚fachlichen Moderation‘,⁹ die von Einzelnen nur bedingt geleistet werden könnte. Das gilt auch für die abschließende Forderung, das Konzept von ‚Urkundenbüchern‘ angesichts der neuen Formen und Möglichkeiten zu überdenken, bis hin zur Frage, ob nicht zur Vermeidung von Doppelarbeiten die Bearbeitung einzelner Fonds aus den Archiven im Vordergrund stehen sollte, und mit der zutreffenden Feststellung, dass die dauerhafte Betreuung des entstandenen ‚offenen Informationssystems‘ nur durch eine institutionelle Anbindung gesichert werden kann.¹⁰

Diese Überlegungen haben – ungeachtet der stürmischen Entwicklungen in diesem Feld – in den letzten sechs Jahren nichts an ihrer Aktualität verloren. Einige ergänzende Überlegungen dazu hat zunächst – in mehreren Beiträgen – Stuart Jenks formuliert, zusammen mit konkreten Umsetzungsvorschlägen.¹¹ Seine Vorschläge laufen auf ein Urkundenbuch als ‚work in progress‘ hinaus, dessen Arbeitsergebnisse – vorläufig ohne Anspruch auf Vollständigkeit und Erfüllung aller editorischen Standards –

⁹ So Sahle/Schaßan, wie Anm. 7, S. 145 und 150.

¹⁰ Ebd., S. 154.

¹¹ Siehe unter anderem Stuart Jenks: Das digitale Preußische Urkundenbuch. In: Beiträge zur Geschichte Westpreußens 17 (2000). S. 181–191. – Ders.: KISS (Keep It Simple, Stupid). Elektronische Quelleneditionen mit einfachsten Mitteln. In: Quellen und Quellenedition im neuen Medienzeitalter. Hg. von Stuart Jenks und Felicitas Schmieder (Fundus; Forum für Geschichte und Ihre Quellen 2). Göttingen 2002. S. 33–43. – Zuletzt (allerdings auf dem Stand von 2001) ders., Diana Kapfenberger und Christina Link: Technischer Fortschritt vs. editorischer Rückschritt. Eine falsche Alternative. In: Hansische Geschichtsblätter 122 (2004). S. 149–162.

sukzessive im Netz veröffentlicht werden, das heißt der ersten Texterfassung sollen nacheinander – und jeweils klar erkennbar – die Schritte der Identifikation der Orts- und Personennamen, die Verschlagwortung und die diplomatische Diskussion folgen. Wesentlich war ihm dabei ein einfaches Modell, das ohne komplizierte technische Voraussetzungen auskommt. Die von ihm vorgesehene ‚Pyramidenstruktur der Editionstätigkeit‘, die die Textaufnahme an verschiedenen Orten mit der Koordination und editorischen Betreuung an zentraler Stelle verbindet, erfordert einen hohen Grad an Organisation, um eine einheitliche Erfassung zu ermöglichen, das heißt die Vereinheitlichung der Namensformen für die Verschlagwortung, eine Übersicht über die verkürzt zitierte Literatur und klare hierarchische Strukturen für die jeweiligen Seiten. Dies soll durch eine Suchmaschine für komplexe Abfragen ergänzt werden.

Patrick Sahle und Georg Vogeler haben 2003 – bzw. in ihrem 2005 erschienenen Beitrag – über 40 einschlägige Projekte untersucht, die in der *Virtual Library Geschichte* vorgestellt sind, und dabei weitergehende Überlegungen zu ‚Urkundenforschung und Urkundenbüchern‘ angestellt.¹² Ihre Analyse erfolgte auf vier Ebenen: Sie galt der Form der Bereitstellung der Urkunden, ihrer Erschließung, den Nutzungsmöglichkeiten und den Überlegungen zu möglichen Zielgruppen. Ihre Ergebnisse sollen hier nicht im Einzelnen wiederholt werden, doch hat die Untersuchung deutlich gemacht, wie disparat bisher die Gestaltung digitaler Urkundenbücher ausfällt. Das betrifft sowohl die Präsentation der Quellen im Netz, als Bild, Regest und digitaler Volltext, wie auch die Gestaltung der ‚Metadaten‘ und den Kontext aus weiterem Material, das der Erschließung der Quellen dient. Nach der Erhebung der Autoren bietet nur etwa ein Drittel der Projekte neu aufgenommene, nicht schon zuvor gedruckte Texte – mit allen Problemen, die dies aufwirft. Sahle und Vogeler schlagen angesichts der wachsenden Unübersichtlichkeit im Bereich der digitalen Editionen ein Urkundenportal vor, das an das Angebot der *Virtual Library Geschichte* anknüpfen, langfristig die angebotenen Ressourcen miteinander verbinden und formal annähern könnte. Ein wichtiges Ziel wäre die Durchsetzung gemeinschaftlicher Standards, um über das Portal auch die Metadaten gemeinsam verwalten zu können.

Im Folgenden seien einige Beispiele für die Umsetzung von Urkundenbüchern im Internet und anderen neuen Medien angeführt und knapp in

¹² Sahle/Vogeler, wie Anm. 3.

Beziehung zu den oben formulierten (theoretischen) Erwartungen gesetzt.¹³

Eine zumeist nützliche, wenn auch einfache Form der digitalen Präsentation von Urkundenbüchern ist die Bereitstellung von Scans, wie sie zum Beispiel seit 1997 bei der Rheinischen Landesbibliothek Koblenz für das 1860 erschienene *Mittelrheinische Urkundenbuch* vorliegen.¹⁴ Die Navigation erfolgt über eine Kopfleiste, einen Überblick geben nur die ebenfalls im Scan vorliegenden herkömmlichen Register der Personen, Orte und Sachen. Ähnlich angelegt ist die digitale Version des *Codex Diplomaticus Saxoniae Regiae*.¹⁵ Hier startet man von der Übersicht über die 25 Bände der Reihe und ruft im einzelnen Urkundenbuch über eine Navigationsleiste die Stücknummern auf; die Erschließung erfolgt nur über die gedruckte Urkundenübersicht und das Register. In beiden Fällen liegt der Vorteil klar darin, ältere Quellenwerke, die selbst in Bibliotheken nicht (oder nur unvollständig) vorhanden sind, vom häuslichen Schreibtisch aus nutzen zu können. Der Vorteil wird allerdings durch die mangelnde Bildqualität dieser und ähnlicher digitaler Urkundenbücher reduziert; im ersten Fall ist die Grenze der Lesbarkeit teilweise erreicht. Dazu kommt, dass die digitalen Möglichkeiten zur Erschließung und Kommentierung der Urkunden nicht genutzt sind.

¹³ Hier kann schon aus Zeitgründen kein Anspruch auf Vollständigkeit erhoben werden; vgl. wiederum die Übersicht unter Adresse: <http://www.vl-ghw.uni-muenchen.de/chartularia.html> (letzte Einsichtnahme am 25.03.2006).

¹⁴ *Urkundenbuch zur Geschichte der, jetzt die preussischen Regierungsbezirke Coblenz und Trier bildenden, mittelrheinischen Territorien*, Bd. 1: Von den ältesten Zeiten bis zum Jahre 1169. Hg. von Heinrich Beyer. Koblenz 1860. Adresse: <http://www.rlb.de/mrHist/91-3152-som/1.1.html> (letzte Einsichtnahme am 25.03.2006).

¹⁵ Adresse: <http://isgv.serveftp.org/codex/> (letzte Einsichtnahme am 27.03.2006). – Vgl. auch die Digitalisierung des *Urkundenbuchs des Hochstifts Merseburg*, 1. Theil: 962–1357. Hg. von Paul Fridolin Kehr. Halle 1899. Adresse: <http://141.84.81.24/cgi-bin/MersUB.pl?seite=I> (letzte Einsichtnahme am 26.03.2006), Teil eines Projekts zur *Merseburger Schriftlichkeit*, das unter anderem ein bisher nur handschriftlich vorliegendes Verzeichnis der Handschriften der Dombibliothek bietet.

CODEX DIPLOMATICUS SAXONIAE REGIAE



Zu den Urkundenbüchern

CDS I A 1 - Urkunden der Markgrafen von Meissen 948-1099
 CDS I A 2 - Urkunden der Markgrafen von Meissen 1100-1195
 CDS I A 3 - Urkunden der Markgrafen von Meissen 1196-1234
 CDS I B 1 - Urkunden der Markgrafen von Meissen 1381-1395
 CDS I B 2 - Urkunden der Markgrafen von Meissen 1396-1406
 CDS I B 3 - Urkunden der Markgrafen von Meissen 1407-1418
 CDS I B 4 - Urkunden der Markgrafen von Meissen 1419-1427

CDS II 1 - Urkunden des Hochstifts Meissen I 963-1356
 CDS II 2 - Urkunden des Hochstifts Meissen II 1357-1423
 CDS II 3 - Urkunden des Hochstifts Meissen III 1423-1492
 CDS II 4 - Urkundenbuch der Stadt Meissen und ihre Klöster
 CDS II 5 - Urkundenbuch der Städte Dresden und Pama

CDS II 6 - Urkundenbuch der Stadt Chemnitz und ihrer Klöster
 CDS II 7 - Urkundenbuch der Städte Kamenz und Löbau
 CDS II 8 - Urkundenbuch der Stadt Leipzig I
 CDS II 9 - Urkundenbuch der Stadt Leipzig II
 CDS II 10 - Urkundenbuch der Stadt Leipzig III
 CDS II 11 - Universität-Leipzig-1409-1555
 CDS II 12 - Urkunden der Stadt Freiberg I
 CDS II 13 - Urkunden der Stadt Freiberg II
 CDS II 14 - Urkunden der Stadt Freiberg III
 CDS II 15 - Urkundenbuch der Stadt Grimma und Klosters Nunsheben
 CDS II 16 - Die Matrikel der Universität Leipzig I
 CDS II 17 - Die Matrikel der Universität Leipzig II
 CDS II 18 - Die Matrikel der Universität Leipzig III / Register

[Zur älteren Version](#)

Projekt Hilfe Kontakt Impressum

Abbildung 1: Startseite des Codex Diplomaticus Saxoniae Regiae.

TEXTS

- [Origo gentis Langobardorum](#)
- [Paulus Diaconus, Historia Langobardorum](#)
- [Leges Langobardorum](#)
- [Historia Langobardorum Codicis Gothani](#)
- [Erchempert, Historiola](#)
- [Andreas Berгамatis, Chronicon](#)
- [Chronica S. Benedicti Casertensis](#)
- [Chronicon Salernitanum](#)
- LANGOBARDISCHE URKUNDEN
- [Codex Cavensis Diplomaticus](#)
- Quellen zur Langobardengeschichte / Sources of Lombard History

LANGOBARDISCHE URKUNDEN (CDL 1)

[EINLEITUNG CDL 1](#) [Originalurkunden](#) [Kopial überlieferte Urkunden](#) [Gefälschte oder verfälschte Urkunden](#) [Der Fälschungskomplex von Cremona](#)

[Einleitung CDL 2](#) [CODICE DIPLOMATICO LONGOBARDO \(2\)](#)

EINLEITUNG (CDL 1)

Edition: *Codice Diplomatico Longobardo* (CDL) 1, ed. L. Schiaparelli (Roma 1929).

Die langobardischen Urkunden bis 774 wurden in den fünf Bänden des *Codice Diplomatico Longobardo* ediert, die mit der Herausgabe der Urkunden von Benevent durch Herbert Zielinski (erscheint voraussichtlich 2003) abgeschlossen sein werden.

Hier sind als Findexilfe die Texte der ältesten Privaturkunden (bis 757) elektronisch erfasst. Für den Apparat und für die exakten Zitate bleibt der Benutzer auf den CDL-Band verwiesen. Abweichend vom CDL wurden die Urkunden hier in Originale, Kopien und Fälschungen aufgeteilt, wobei der (vom Barockgelehrten Dragoni angefertigte) Fälschungskomplex von Cremona eigens zusammengefaßt wurde. Grundlage der Einteilung war die Edition. Bei kopialer Überlieferung, besonders wenn sie zeitfremd ist, muß darüber hinaus die Möglichkeit einer Verfälschung über das vom Editor angenommene Maß hinaus in Betracht gezogen werden. Die Originalurkunden sind inzwischen als Faksimile mit Transkription, ausführlichem Apparat und Übersetzung in den betreffenden Bänden der *Chartae Latinae Antiquiores* herausgegeben worden, deren Benützung sich sehr empfiehlt.

Ausgewählte Literatur:

- *Codice Diplomatico Longobardo*, Bd. 1-3, hrg. L. Schiaparelli, 1929-33; Bd. 3 u. 4, hrg. C. Brühl, 1973-81; Bd. 5, hrg. H. Zielinski, 1986.
- A. Castagnetti, *L'organizzazione del territorio rurale nel medioevo*, 1979.
- M. Cavina, *Sculdahis rector loci et minor iudex*, *Rivista della Storia del Diritto Italiano* 65, 1992, 145-68.
- P. M. Coni, *La Tuscia e i suoi ornamenti territoriali nell'alto medio ev.* In: *Atti del V Congresso internazionale di studi sull'alto medioevo* (Spoleto 1973) 61-116.

Abbildung 2: Startseite der „Langobardischen Urkunden“.

Einen Schritt weiter geht ein privates Vorhaben, die Digitalisierung des 6. Bandes des *Oldenburgischen Urkundenbuches für Jever und Knipphausen*.¹⁶ Sie präsentiert die Quellen in der Reihenfolge der Vorlage im HTML-Format, verlinkt im Register die Stücknummern mit dem Text und erlaubt eine einfache Suche nach Nummern und Begriffen. In gewisser Hinsicht vergleichbar sind die in einem weiteren Kontext von Quellen zur langobardischen Geschichte angebotenen *Langobardischen Urkunden*, die für die Annotierung ausdrücklich auf die gedruckte Edition von Luigi Schiaparelli verweisen.¹⁷ Die Stücke sind nach den Nummern der Edition in wenigen HTML-Dateien angeordnet, allerdings getrennt nach Originalurkunden, Kopialüberlieferung und Fälschungen. Beiden digitalen Urkundenbüchern gemeinsam ist die Digitalisierung gedruckten Urkundenmaterials, doch ist die Suche weiterhin auf die Nummern der Druckausgabe konzentriert; eine Suche nach Namen oder Begriffen ist nur über den Browser möglich. Zudem verzichten gerade die *Langobardischen Urkunden* auf die in der Druckfassung enthaltenen Zusatzinformationen.

Weit mehr verspricht die Digitalisierung des *Corpus Altdeutscher Originalurkunden bis 1300* mit rund 4 400 Stücken, von dem bereits eine – allerdings eingeschränkte – Testversion für das Rhein-Maas-Gebiet online verfügbar ist.¹⁸ Die digitale Textfassung soll nach internationalen Standards ausgezeichnet werden und in das *Mittelhochdeutsche Textarchiv* eingehen, wird also in einen größeren Kontext integriert.¹⁹

¹⁶ Startseite: <http://www.urkundenbuch.net/index.php> (letzte Einsichtnahme am 27.03.2006); basierend auf *Oldenburgisches Urkundenbuch, VI: Urkundenbuch von Jever und Knipphausen*. Hg. von Gustav Rüthing. Oldenburg 1932, initiiert durch ein Antiquariat.

¹⁷ Adresse: http://www.oew.ac.at/gema/lango_urk.htm (letzte Einsichtnahme am 27.03.2006); basierend auf *Codice Diplomatico Longobardo*. Bd. 1–2. Hg. von Luigi Schiaparelli. Rom 1929–1933.

¹⁸ Startseite unter Adresse: <http://germazope.uni-trier.de/Projects/Corpus/start/index.html> (letzte Einsichtnahme am 27.03.2006); die Testversion für das Rhein-Maas-Gebiet unter <http://gepc189.uni-trier.de/cgi-bin/iRMnet/RMnetIndex.tcl?hea=qf&for=qfcoraltdu>; digitalisiert wird das abgeschlossene *Corpus der altdeutschen Originalurkunden bis zum Jahr 1300*, Bd. I: 1200–1282. Hg. von Friedrich Wilhelm; Bd. II: 1283–1292. Hg. von Friedrich Wilhelm und Richard Newald; Bd. III: 1293–1296. Hg. von Richard Newald, Helmut de Boor und Diether Haacke. Bd. IV: 1297–(Ende 13. Jahrhundert). Hg. von Helmut de Boor und Diether Haacke. Bd. V: Nachträge. Hg. von Helmut de Boor, Diether Haacke und Bettina Kirschstein. Lahr 1932–1986.

¹⁹ Die gedruckte Ausgabe des *Corpus altdeutscher Originalurkunden* wird mit etwas anderen Zielen und in räumlicher Beschränkung, durch das 1990 begonnene *Trierer Korpus mittelfränkischer Urkunden bis 1330* fortgeführt, Adresse: <http://136.199.110.189/cgi-bin/RMnetIndex.tcl?hea=qf&for=qmfranku> (letzte Einsichtnahme am 13.02.2005), das zurzeit 324 Urkunden umfasst; zur älteren Version des Projekts vgl. die Beiträge von Andrea Rapp: Computergestützte

Quellenforum <RM.net>

Forschungsforum Themenforum Kartenforum Quellenforum

Sie sind hier: Urkunden des Rhein-Maas-Raumes im Rahmen des Corpus der altdeutschen Originalurkunden

Dokument

voriger Datensatz << >> nächster Datensatz

2. 1227.

Reg. 1.05 Das sogenannte Ottonische Stadtrecht der Stadt Braunschweig. Dreispaltig geschrieben. — —

1.2.001.12 Swelich^{tt} voget enen richtere fet an fine ftat · fwaz vor dheme gelent wert · dat fal

1.2.001.13 gelike ftede wesen · alfe it de voget felue ftedegede.

1.2.001.14 Swelich man deme anderen fculdich es · vn̄ es ime vorfaketh · entgeit he if ime met

1.2.001.15 tvge · oder met fime êthe · he ne heuet weder dat gerichte nicht verloren wane dat gelt

1.2.001.16 al ene.

1.2.001.17 Swelich man sich finef tvgef beropet vmbe gelt · vn̄ if ime borft wert · he ne darf

1.2.001.18 dheme richte nicht wedden wane ver fcill.

1.2.001.19 Swelich man den anderen w^vndit ove dot fleit · vn̄ vlu^ochtich wert · heuet

Abbildung 3: Testversion des „Corpus der altdeutschen Originalurkunden“.

Bisher kann wiederum nur nach Nummern gesucht bzw. zwischen den einzelnen Datensätzen ‚geblättert‘ werden. Das sollte sich aber nach Abschluss der Arbeiten am Corpus ändern. Da die Edition literaturwissenschaftlichen Standards genügen muss, ist auf eine Besonderheit hinzuweisen, die von allgemeiner Bedeutung ist: Sie bietet eine diplomatische Edition ohne die ‚Eingriffe‘ geschichtswissenschaftlicher Editoren und gibt auch graphische Zeichen im Text wieder. In der Tat sollten die digitalen Editionen mittelfristig nicht nur den Anforderungen einer Wissenschaft nachkommen, sondern auch Optionen für andere methodische Zugänge eröffnen.

Verfahren zur Erfassung, Katalogisierung, Bearbeitung und Edition mittelfränkischer Urkunden des 13. und 14. Jahrhunderts. In: Maschinelle Verarbeitung altdeutscher Texte V. Hg. von Stephan Moser, Peter Stahl, Werner Wegstein und Norbert Richard Wolf. Tübingen 2001. S. 247–261. – Dies.: Die elektronische Edition, Erschließung und Vernetzung des *Trierer Korpus mittelfränkischer Urkunden des 14. Jahrhunderts*. In: Jahrbuch für Computerphilologie 2000. S. 147–161. – Dies.: Die elektronische Edition, Erschließung und Vernetzung historischer Quellen und Grundlagenwerke. Das *Trierer Korpus mittelfränkischer Urkunden des 14. Jahrhunderts*. In: TUSTEP educa. Actas de congreso des International TUSTEP User Group. Hg. von Nicolás Castrillo Benito und Peter Stahl. Burgos 2001. S. 99–123.

The screenshot shows a web interface for a virtual archive. On the left is a navigation menu for 'Mönasterium' with sub-categories for Austria (NÖ, OÖ, Sbg, T), Czech Republic, Germany, Hungary, and Slovenia. The main content area displays details for 'Mönasterium Teikonik' (AT Österreich, NÖe Niederösterreich) with a list of holdings including 'Heiligenkreuz, OCist - Heiligenkreuz, Zisterzienser (1136-1399)'. Below this is a 'Beschreibung' section and a 'Verzeichnungseinheiten:' table. At the bottom, there is a search results summary with a grid of links for document ranges.

Lfd.-Num.	Titel	Datierung
1/051	Der päpstliche Cardinal-Legat Conrad bestätigt der Abtei heiligen Kreuz das Patronat über die Kirche zu Sulz.	1225. April 1. Heiligen Kreuz.
1/052	Der päpstliche Cardinal-Legat Conrad bestätigt der Abtei heiligen Kreuz den unter Vermittelung des Herzogs Leopold von Österreich mit der Abtei Melk abgeschlossenen Vergleich wegen der Zehente.	1225. April 1. Heiligen Kreuz.
	Benedict Hinnricus III bestätigt der Abtei heiligen Kreuz das Patronat	1297 Januar

Ergebnisse: [c-\[1-10\]](#) [\[11-20\]](#) [\[21-30\]](#) [\[31-40\]](#) [\[41-50\]](#) [\[51-60\]](#) [\[61-70\]](#) [\[71-80\]](#) [\[81-90\]](#) [\[91-100\]](#) [\[101-110\]](#) [\[111-120\]](#) [\[121-130\]](#) [\[131-140\]](#) [\[141-150\]](#) [\[151-160\]](#) [\[161-170\]](#) [\[171-180\]](#) [\[181-190\]](#) [\[191-200\]](#) [\[201-210\]](#) [\[211-220\]](#) [\[221-230\]](#) [\[231-240\]](#) [\[241-250\]](#) [\[251-260\]](#) [\[261-270\]](#) [\[271-280\]](#)

Abbildung 4: Das „virtuelle Archiv mitteleuropäischer Klöster und Bistümer“.

Eine komplette ‚Neuschöpfung‘ ist das *Virtuelle Archiv mitteleuropäischer Klöster und Bistümer*,²⁰ das ein Urkundenbuch als ‚work in progress‘ darstellt, als Gemeinschaftsprojekt niederösterreichischer Archive begann, nunmehr aber mehrere Regionen erfasst. Es verbindet Archivfindmittel, Inhaltsangaben, Beschreibungen, einige Volltexte und Abbildungen mit einer Reihe weiterer Hilfsmittel wie einem Glossar und mit Links zu weiteren Quellen im Netz. Die Urkunden der bisher erfassten Klöster lassen sich über eine zeitlich eingrenzbare Suche erschließen und die allerdings nicht immer ganz einfache Navigation erfolgt über eine Mehrfensterlösung.²¹ Der besondere Reiz dieser Urkundenpräsentation ist die Zusammenarbeit der Archive, die so ihre Bestände besser zugänglich machen.

Die drei in Hamburg betriebenen Urkundenbücher bieten anders als die zuletzt angesprochenen Projekte sowohl bereits gedruckte Quellen – aus anderen Urkundenbüchern, Quellensammlungen und Anhängen in Monographien und Aufsätzen – als auch laufend neues Material aus den Archiven, das erschlossen und ediert wird. Das virtuelle *Preußische Urkundenbuch*, das

²⁰ Adresse: <http://www.monasterium.net/> (letzte Einsichtnahme am 27.03.2006).

²¹ Indexseite: http://www.mom.findbuch.net/php/main.php?ar_id=3263&action=rech (letzte Einsichtnahme am 27.03.2006).

das gedruckte Urkundenbuch für die Jahre 1382 bis 1525 fortsetzen soll,²² umfasst inzwischen rund 6 100 Stücke (darunter knapp 5 000 digitalisierte Regesten aus einem gedruckten Archivrepertorium,²³ aber auch schon mehrere Hundert Volltexte). Das virtuelle *Hamburgische Urkundenbuch* für die Jahre 1351 bis 1529 enthält rund 220,²⁴ die *Hospitaller Sources* bieten für den Zeitraum 1310 bis 1522 rund 230 Stücke,²⁵ jeweils mit mehr als 50 Prozent Volltexten, mit sehr unterschiedlichem, jeweils ausgewiesenem Bearbeitungsstand. Der Einstieg dazu erfolgt über Jahresregestenlisten, die Erschließung bietet teilweise eine einfache Suchmaschine, sofern nicht die Browser-Funktionen genutzt werden müssen. Alle drei Urkundenbuchprojekte sind inzwischen in einen größeren Kontext von eigenen und fremden Angeboten eingebunden, der auch Scans älterer Drucke und Bildmaterial einschließt.

Dieser naturgemäß begrenzte Überblick über digitale Urkundenbücher hat zunächst einmal den Eindruck der Vielfalt der Zugänge verstärkt. Tatsächlich erfüllen Sie zumindest einen Teil der eingangs formulierten Anforderungen.

- Die digitalen Urkundenbücher machen – dort, wo nicht nur Bilder geboten werden – die Quellen besser nutzbar und verfügbar;
- sie vermitteln mit Ausnahmen auch das jenseits der Urkundentexte im Druck ‚gespeicherte‘, erschließende Wissen;
- sie eröffnen die Möglichkeit zur Vernetzung mit weiteren Quellen und Informationen und sind weithin ergänzbar;
- die Erarbeitung digitaler Editionen hat inzwischen einen Diskussionsprozess über die künftige Gestaltung dieser Angebote angestoßen.

Deutlich geworden ist allerdings auch, dass es einer stärkeren Standardisierung bedarf, sowohl in der Auszeichnung der Texte, damit eine bessere Er-

²² Adresse: <http://www1.uni-hamburg.de/Landesforschung/orden.html> (letzte Einsichtnahme am 27.03.2006).

²³ Erich Joachim (Bearbeiter): *Regesta Historico-Diplomatica Ordinis Sanctae Mariae Theutonicorum, 1198–1525*. Hg. von Walther Hubatsch. Pars I, Bände 1–3, Pars II, Registerband zu I–II. Göttingen 1948–1973, verzeichnet Bestände der XX. Hauptabteilung des Geheimen Staatsarchivs Preussischer Kulturbesitz, Berlin (Historisches Staatsarchiv Königsberg).

²⁴ Adresse: http://www1.uni-hamburg.de/hamburgisches_ub/HambUB.html (letzte Einsichtnahme am 27.03.2006).

²⁵ Adresse: <http://www1.uni-hamburg.de/hospitallers/hospitallers.html> (letzte Einsichtnahme am 27.03.2006).

schließung durch die Suchinstrumente erfolgen kann, als auch in den ‚Mindest‘-Anforderungen an die Urkundenbücher. So sollten Bilder eine gute Lesbarkeit garantieren; Scans der Originale müssten idealerweise sogar in mehreren Stufen und ‚Auflösungen‘ verfügbar gemacht werden; ein ‚Mindest‘-Satz von Metadaten sollte geboten werden. Das kann leicht mit dem von Patrick Sahle und Georg Vogeler vorgeschlagenen Urkundenportal in Verbindung gebracht werden: Ein solches Portal, das vielleicht in der Tat von der entsprechenden Seite der *Virtual Library Geschichte* seinen Ausgang nehmen könnte, würde eine solche Standardisierung fördern. Ein denkbare Nebeneffekt ist auch die Vermeidung von Doppelangeboten. Dieselbe Quelle müsste nicht zweimal, etwa für das *Hansische* und das *Preußische Urkundenbuch*, aufgenommen werden, sondern könnte in beide unter denselben Metadaten integriert werden.

Diese Überlegung zieht unweigerlich die Frage nach der Finanzierung eines solchen Portals nach sich. Im Hamburger Kontext haben sich bei der Bearbeitung des Quellenmaterials grundlegende Probleme gezeigt. Die drei Projekte sind dem von Stuart Jenks entwickelten Modell eines Urkundenbuchs als ‚work in progress‘ verpflichtet, das auf der Kooperation zwischen mehreren Wissenschaftlerinnen und Wissenschaftlern, aber auch auf der Integration von Studierenden aufbaut, die über Lehrveranstaltungen und Abschlussarbeiten eingebunden werden sollen.²⁶ Es hat sich aber gezeigt, dass dieser Ansatz nur sehr begrenzt funktioniert. Der Appell an andere Wissenschaftlerinnen und Wissenschaftler, die von ihnen digitalisierten Quellen in dafür vorliegenden ‚Formularen‘ zur Verfügung zu stellen, hat bisher nur wenig Resonanz gefunden²⁷ und es sind zumeist nur wenige Dateien dazugekommen.²⁸ Ähnliches gilt für die Einbindung von Studierenden. Selbst innerhalb eines zweijährigen E-Learning-Projekts wurden so nur relativ wenige Quellen aufgenommen.²⁹ Den Möglichkeiten eines ‚Her-

²⁶ Allgemein dazu wie Anm. 11.

²⁷ Mit Ausnahme der Zusammenarbeit mit Dieter Heckmann und Stuart Jenks (für das *Preußische Urkundenbuch*) sowie mit Klaus-Joachim Lorenzen-Schmidt (für das *Hamburgische Urkundenbuch*).

²⁸ Zu nennen sind hier weiter Arno Mentzel-Reuters, Jan-Erik Beuttel, Marc Schmidt, Paweł Jeziorski und Krzysztof Kopinski (für das *Preußische Urkundenbuch*) sowie Karl Borchardt (für die *Hospitaller Sources*).

²⁹ Im Rahmen des E-Learning-Projekts *Geisteswissenschaften im Internet*, Adresse: <http://www.phil-gesch.uni-hamburg.de/fbelch.html> (letzte Einsichtnahme am 15.02.2005), strebte das Teilprojekt *Edition mittelalterlicher Quellen im Internet* die Heranführung Studierender ans wissenschaftli-

ausgebers' sind dabei klare Grenzen gesetzt. Nicht zufällig wurde die von Patrick Sahle, Thorsten Schaßan und Stuart Jenks vorgeschlagene und in der virtuellen Erlanger Fassung des gedruckten *Preußischen Urkundenbuchs* durch ein Eingabeformular realisierte interaktive Verschlagwortung durch die Benutzer völlig aufgegeben.³⁰ Schon die rein technische Überprüfung der in Kooperation bearbeiteten Dateien erfordert einigen Zeitaufwand; eine komplette formale Angleichung, die eine Voraussetzung für die Integration in ein Urkundenportal wäre, wäre eine ‚hauptamtliche‘ Aufgabe, von der Betreuung des Portals ganz zu schweigen. ‚Nebenamtliche‘ Betreuer wären somit überfordert, man müsste ein bundesweites oder über Deutschland hinausreichendes Projekt beantragen, ähnlich wie für das *Virtuelle Archiv mitteleuropäischer Bistümer und Klöster*, das aber auch eine langfristige Pflege ermöglicht.

Ingrid Heidrich hat 2004 in einer Rezension des virtuellen *Hamburgischen Urkundenbuchs* unter anderem die wenig spezifische Suchfunktion und die Zersplitterung des Materials kritisiert.³¹ Die Suchfunktionen sind in der Tat ein wichtiges Element, das bisher – wie bei vielen Projekten – unzureichend ausgebildet ist; während die Disparatheit des in den drei Ham-

che Arbeiten und an die digitale Quellenedition an. Die wenigen größeren Projekte, die daraus erwachsen sind, sind noch nicht abgeschlossen. Allerdings bieten die Dateien von Sebastian Kubon und Martin Neumann (für das *Preußische Urkundenbuch*), Mathias Nagel, Annika Souhr, Jeanine Marquard und Susanne Dreyer (für das *Hamburgische Urkundenbuch*) sowie von Manuela Plener, Maike Claußnitzer und Marcus Prey (für die *Hospitaller Sources*) jeweils wichtige Ergänzungen (neben weiteren ‚Vorarbeiten‘ aus Seminaren). Größere Beiträge für das *Preußische Urkundenbuch* wurden zudem durch Hilfskräfte erarbeitet (Frauke Schmitz, Jorun Poettering, Joachim Laczny und Markus Kompisch).

³⁰ Dazu unter anderem Jenks und andere, Fortschritt, wie Anm. 11, S. 159. – Das Formular generierte eine E-Mail an Stuart Jenks, der dann sein Stichwortregister, das er für die früheren, aus dem gedruckten Urkundenbuch stammenden Teile angelegt hat, entsprechend ergänzt hätte. Dieses Tool wurde jedoch nie ernsthaft genutzt – ein Argument für die ‚Filterung‘ durch einen Bearbeiter –, andernfalls aber wäre der Arbeitsaufwand erheblich gewesen.

³¹ Veröffentlicht unter der Adresse: <http://hsozkult.geschichte.hu-berlin.de/rezensionen/id=64&type=rezwww> (letzte Einsichtnahme am 05.11.2004; 28.03.2006). – Kleinere Kritikpunkte Heidrichs sind durchaus diskussionswürdig. Die Frage, ob ein ‚Impressum‘ (im Netz ohnehin etwas ungewöhnlich) erforderlich ist, ist meines Erachtens sekundär, wenn die Betreuer und Bearbeiter des Projekts mit Kontaktadressen genannt sind. Dagegen sollte jede Einzeldatei eindeutig die Identifikation als Teil des Gesamtprojekts erlauben, da sie ja über Suchmaschinen einzeln erschließbar sein sollte. Der Verzicht auf eine Kopfzeile „Das virtuelle Hamburgische Urkundenbuch“ ist daher nicht denkbar, auch wenn natürlich die graphische Gestaltung immer Frage des Geschmacks ist. Allerdings sollte eine gute Lesbarkeit gewährleistet sein (was vielleicht heute nicht immer der Fall ist), ebenso ohne zuviel Scrollen.

burger Projekten bisher dargebotenen Materials eine Folge des kooperativen Vorgehens darstellt, das auf Masse setzt, die nur – und das auch nur mit Einschränkungen – im Fall des *Preußischen Urkundenbuchs* erreicht ist. Ein Urkundenportal sähe sich beiden Problemen in höherer Potenz gegenüber. Der Zugriff der Suchinstrumente müsste auf das gesamte, auf verschiedenen Servern liegende Material zielen, aber auch spezifizierte Einzelabfragen ermöglichen, und anders als beim *Virtuellen Archiv mitteleuropäischer Bistümer und Klöster* fehlte eine inhaltliche Klammer. Dennoch sollte der Versuch unternommen werden, ein derartiges Urkundenportal ins Leben zu rufen und die genannten Schwierigkeiten zu überwinden.

Das forschende Mitglied der E-Humanities, der mit digitalen Medien arbeitenden Geisteswissenschaften, wird aus heutiger Perspektive wahrscheinlich mittelfristig eine Reihe von Erwartungen an die digitalen Urkundenbücher zur mittelalterlichen Geschichte richten, die nur durch verstärkte Diskussion und Kooperation – und letztlich über ein Urkundenportal – zu erreichen sind:

- Wie schon heute, wird auch künftig in der digitalen Welt vor allem die Schnelligkeit des Zugriffs zählen. Das ist bereits der Vorteil der als Bild digitalisierten Publikationen, die bei aller Problematik die – nicht immer leichte – Beschaffung über die Bibliotheken ersetzt; umso besser ist, beim digitalen Zugriff nicht mehr nur ein Urkundenbuch, sondern gleich mehrere – oder möglichst viele – mit den eigenen Fragestellungen durchzusuchen zu können. Der so erreichbare Fortschritt ist heute bereits im Bereich der Handschriften erkennbar, für die die Seite der *Manuskripta Mediaevalia* über 40 000 Manuskripte erschließt.³²
- Die Nutzer des Urkundenportals werden eine stärkere Standardisierung erwarten. Dazu gehört sicherlich die einheitliche Auszeichnung der Texte, ohne die die Suchinstrumente unzuverlässige Ergebnisse erzeugen würden, dazu gehört aber auch eine technische Annäherung und ein vergleichbarer Stand der Hilfsmittel-Angebote. Gerade in letzterer Hinsicht kann ein Portal Erleichterungen bringen, durch die Kooperation zwischen verschiedenen Projekten.
- Schließlich werden die Nutzer hoffen, dass es zu einer Verdichtung und Verbesserung des Angebots kommt. Die Probleme eines kooperativen

³² Adresse: <http://www.manuscripta-mediaevalia.de/> (letzte Einsichtnahme am 28.03.2006). – Dort haben allerdings urheberrechtliche Probleme zur Reduktion des Angebots geführt.

Vorgehens ohne tragfähige Projektförderung habe ich gerade am Beispiel der Hamburger Urkundenbuchprojekte verdeutlicht. Die – möglichst europaweite – Zusammenarbeit zwischen Urkundenbuchprojekten, auch in Gestalt eines gemeinsamen Portals, könnte auf dreifache Weise im Sinne der angesprochenen Erwartung wirken. Zum einen würde sie eine vertiefte Debatte über die Formen der Bereitstellung von Quellen anstoßen, die auch die Nutzer einbezieht; zum anderen könnten gemeinsame Werkzeuge entstehen, die die kooperative Arbeit an den Projekten erleichtert, zum Beispiel so etwas wie ein ‚Autorentool‘ für die beteiligten Wissenschaftlerinnen und Wissenschaftler; und schließlich ließe sich gemeinsam – auch mit interessierten Archiven – vielleicht mit größerer Aussicht auf Erfolg eine Finanzierung für das Projekt finden.

Schon heute ist Forschung außerhalb der digitalen Welt kaum mehr – oder vielleicht schon: nicht mehr – möglich, die Tendenz wird sich weiter verstärken. Die wachsende Vielfalt von Quellenangeboten im World Wide Web bedarf auch der Qualitätssicherung, die nur durch Zusammenarbeit erreichbar ist. Die Diskussion über die digitalen Urkundenbücher zur mittelalterlichen Geschichte kann einen wichtigen Beitrag dazu leisten.

Verborgen, vergessen, verloren?

Perspektiven der Quellenerschließung durch die digitalen *Regesta Imperii*¹

Dieter Rübsamen und Andreas Kuczera

Die *Regesta Imperii* präsentieren sich heute – wie schon seit Jahrzehnten – als eines der meistgenutzten mediävistischen Grundlagenwerke zur europäischen Geschichte des Mittelalters. Das von dem Frankfurter Stadtbibliothekar Johann Friedrich Böhmer im ersten Viertel des 19. Jahrhunderts initiierte und zunächst weitgehend im Alleingang betriebene Unternehmen hatte als vorderstes Ziel, zu jedem Herrscher, worunter die römisch-deutschen Kaiser und Könige sowie für etliche Epochen auch die Päpste zählen, eine lückenlose Dokumentation der jeweiligen Person(en) und ihres politischen Handelns vorzulegen. Diese Herrscherregesten wurden – zumal in den älteren Bänden – ergänzt von ‚Reichssachen‘, worunter zumeist urkundliches Material, das in irgendeinem inhaltlichen Bezug zur Reichsgewalt steht, zu verstehen ist. Zur Erstellung eines im Idealfall vollständigen Itinerars, also eines lückenlosen Nachweises der Aufenthaltsorte und der dortigen Anwesenheitszeit eines Herrschers, mithin der Reiseweg des noch nicht durch Residenzen örtlich verfestigten Königtums, wurden alle schriftlichen Quellen mit entsprechendem Herrscherbezug herangezogen: in erster Linie Urkunden, deren Datierungsangaben eindeutige chronologische Zuordnungen erlauben, daneben aber auch – und gerade für die früheren Epochen wichtig – Historiographisches. Die Böhmersche Grundkonzeption sah dabei durchaus eine Bescheidung auf die dienende Funktion zur Er-

¹ Dem Tagungscharakter entsprechend wird die Vortragsfassung für die Druckversion beibehalten. Anstelle von Literaturangaben zu den *Regesta Imperii* sei verwiesen auf die entsprechenden Verzeichnisse der Homepage <http://www.regesta-imperii.de/> (letzte Einsichtnahme am 10.04.2006).

schließung der zeitgleich im Entstehen begriffenen großen Reihen der *Monumenta Germaniae Historica* – insbesondere der einschlägigen Urkundenreihe der *Diplomata* – vor: Vorarbeit und Hilfsmittel zugleich für die dann endgültigen Editionen. Schon die ersten Nachfolger Böhmers entwickelten jedoch dieses karge Konzept weiter, hin zu dem, wie wir es heute kennen: Ein Regest soll demnach eine möglichst exakte Wiedergabe des Rechtsinhalts von Urkunden, deren formale Beschreibung, die kritische Einordnung der historiographischen Nachrichten bieten, alles streng orientiert an der Chronologie. Dieses Konzept hat sich über Generationen von Forschern hinweg bewährt und wird – ohne jemals verbindlich für alle Unternehmungen festgelegt worden zu sein – fortgeführt werden, auch im digitalen Zeitalter.

Natürlich kommen heutige, neuere Regesten schon allein optisch in anderem Gewand daher als ältere. Das springt in den gedruckten Versionen geradezu ins Auge, in der digitalen aber erst auf den zweiten Blick. Und genau hier wollen wir einhaken und mit durchaus selbstkritischem Unterton die Frage stellen:

Was mag angesichts einer fast 180 Jahre währenden, viele tausende gedruckte Seiten umfassenden Publikationsstrecke nun verborgen, gar vergessen sein oder muß letztlich sogar als verloren gelten? Und was kann diese Fragestellung vor dem Hintergrund unserer Tagungsproblematik leisten? Es geht also hier nicht so sehr darum, das bisher digital Erreichte zu präsentieren, hier mag sich jeder einen schnellen Überblick mit einem Klick auf unsere Homepage verschaffen, sondern vielmehr die durch die digitale Version sich ergebenden Möglichkeiten und Erweiterbarkeiten zu beschreiben und auch die sich daraus ergebenden Probleme in der digitalen Werkstatt anzureißen.

Wir wollen uns von zwei Seiten nähern, von innen nach außen vorgehend. Am Ende mag dann das eine oder andere Fragezeichen wegfallen, vielleicht kommen aber auch durch die sich hoffentlich entzündende Diskussion neue hinzu!

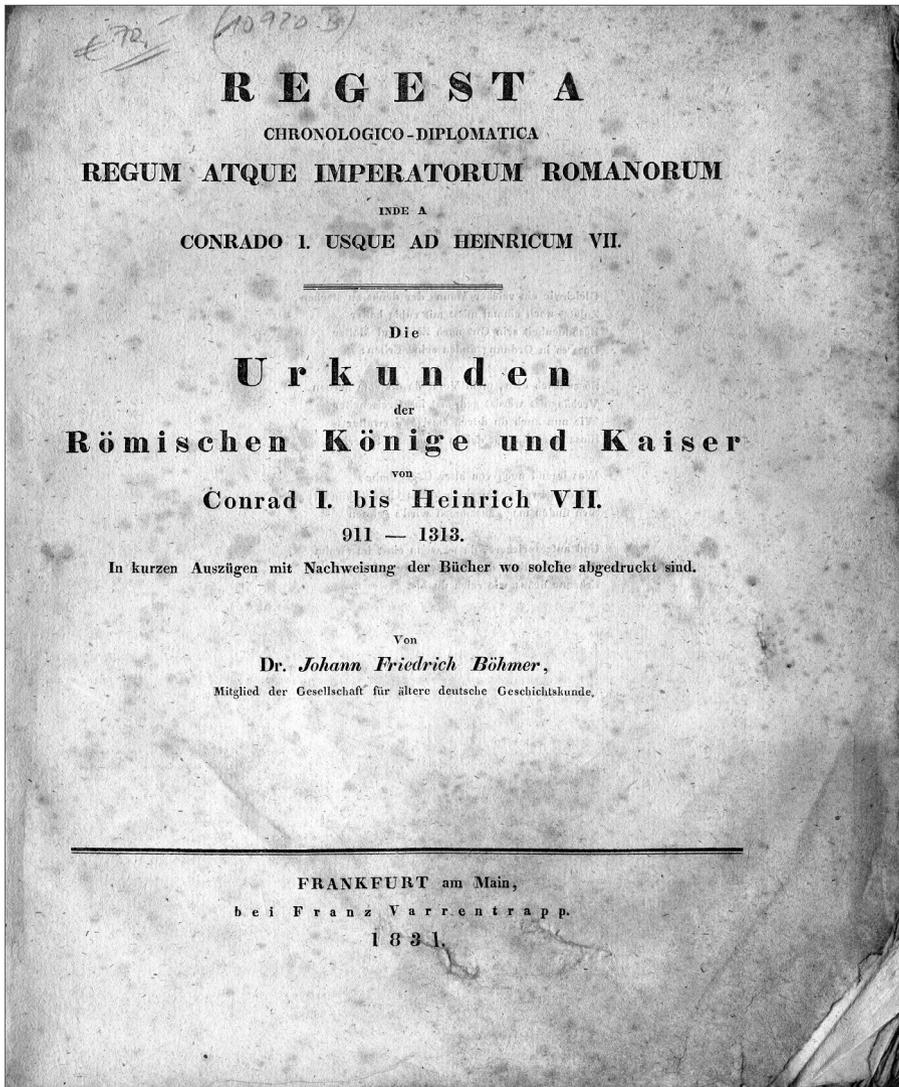


Abbildung 1: Titelblatt des „Ur-Böhmer“.

Bei aller Konzeptionsstabilität ist zunächst einmal festzuhalten, dass das durch die *Regesta Imperii* in ihren vierzehn klassischen Abteilungen – orientiert an Herrscherdynastien, späterhin auch an einzelnen Herrschern – dargebotene Material ein äußerst heterogenes Erscheinungsbild liefert. Ein knapper Überblick mag dies veranschaulichen.

Der *Ur-Böhmer*, sozusagen die Vorabausgabe, die noch nicht in die klassische Zählung aufgenommen wurde, stammt aus dem Jahr 1831. In diesem Band hat Böhmer das gesamte ihm durch die Quellen-Literatur bekannte Material von Konrad I. bis Heinrich VII., immerhin aus einem Zeitraum von 400 Jahren, gesammelt. Und: Selbst diese kargen Schnipsel sind noch heute, mangels bisher erfolgter Neubearbeitungen, in Teilen (zum Beispiel für die das 11. Jahrhundert dominierende Herrscherfamilie der Salier) zitierfähig!

Heinrich III. 1039 — 1040. (VII — VIII).					73
No.	O.	R.	1039 (VII)		
	12	1	Juni 4.	Traiecti	<i>Regierungsantritt.</i> Bereits am 14. April 1028 war er zu Aachen als König gesalbt worden und rechnet von da an die Jahre seiner Königsweihe.
1445	—	—	22.	Anternacho	bestätigt die Privilegien des Bisthums Minden. Schaten I, 515. Pistorius S. S. III, 822.
1446	—	Juli	10.	Mogontiae	bestätigt die Privilegien des Hochstifts Bamberg. Heyberger Nr. 24 mit Siegel.
1447	—	Sept.	3.	Goslariae	bestätigt dem Abt Truchtmar die Privilegien und Besitzungen der Klöster Corvei und Herford. Schaten I, 516. Falke 740.
1448	—	—	3.	—	bestätigt dem Kloster Corvei freie Abtwahl, Immunität, Zehntfreiheit etc. Schaten I, 518.
1449	—	—	3.	—	Schutz-, Immunitäts- und Bestätigungsbrief für das Nonnenkloster zu Kemnata im Gau Tilithi. Schaten I, 519.
1450	—	—	3.	—	bestätigt die Privilegien und Besitzungen des Klosters Gandersheim. Harenberg 672.

Abbildung 2: Regesten Heinrichs III. aus dem *Ur-Böhmer* (1831).

Das bereits angesprochene erweiterte Regestenkonzept der Nachfolger Böhmers mag an einem Auszug aus den *Regesta Imperii V* (Jüngere Staufer), nach altem, an die Bearbeiternamen angelehnten Zitierstatus auch als Böhmer-Ficker-Winkelmann bezeichnet, deutlich werden.

Dieses Druckbild, charakterisiert durch die vertikalen Kolumnen von Datum, (Ausstellungs-)Ort und eigentlichem Regestentext, wurde gegen Ende des 19. Jahrhunderts gleichsam kanonisiert, alle Bände der älteren Bearbeitungen sind in diesem Format erschienen.

aug. 22	Maguntie	An diesem tage geht der kaiser gekrönt und gibt den fürsten ein grosses fest. Ann. Col. — Imperator in octava assumptionis coronatus incesit. Ann. Marbac. Vgl. Sächs. Weltchr. c. 379. — Setzen Albricus und die Chr. Mettensis, M. Germ. 24,521, den hoftag überhaupt auf die octava assumptionis, so haben sie wohl nur diesen hauptfesttag im auge. *a
— 23	—	bestätigt dem domecapitel zu Hildesheim auf bitte des bischofs Conrad von Hildesheim den rückkauf gewisser vogteien welche Hermann und Heinrich gebrüder von Woldenberg und andere von dessen kirche trugen. Harenberg Hist. Gand. 428. Huill. 4,758. 2105
— 23	—	Conrad von Hohenlohe, graf der Romagna, bekundet dass er seine burg Leindal der kirche von Cöln aufgetragen habe und damit vom erzbischof Heinrich belehnt sei. Z.: Heinr. gr. v. Sain, Loth. gr. v. Hochstaden, Gerl. v. Büdingen, Hermann v. Müllenark, Alb. v. Drincht, Rosemann v. Kempenich, Gerl. u. Heinr. br. v. Isenburg, Gerh. vogt v. Cöln, Dietr. truchs., Goswin marschall, Alb. v. Hörde schenk, Winemar v. Asterlo. Hansselmann Landeshoheit 1,401. Isenburg. Geschlechtsreg. 2,45. Wirtemb. Urkkb. 3,363. *2106

Abbildung 3: Auszug aus den RI V (Jüngere Staufer).

Im Unterschied zum *Ur-Böhmer* finden sich nun längere Regestentexte, Angaben von Zeugen, ein diplomatischer Apparat, Literaturangaben und Kommentare. Dies alles nur spärlich strukturiert, aus Sparsamkeitsgründen (Papier!) oft mit Kürzungen aller Art versehen, für den Buchnutzer akzeptabel, für die Digitalisierung, das sei vorweg festgehalten, problematisch. Beispielhaft sei hier ein Ausschnitt aus dem Register der *Regesten Kaisers Sigismunds (RI XI)* gezeigt, der die Problematik verdeutlichen kann.

805. 833 f. 852. 855. 861. 881 f.
 886. 919. 922. 935. 938. 3006.
 025 f. 029. 031 f. 035. 039 ff. 045.
 047 ff. 062A. 103. 107. 114. 118.
 120. 129 f. 139 f. 142. 144. 144^a.
 147. 151^a. 160 f. 223. 243. 309.
 441^a. 481 ff. 495. 599. 600 ff. 688.
 720. 737 f. 778 ff. 782. 795. 812.
 814. 944. 968. 973. 4040. 047. 049.
 155. 390 f. 398. 417 f. 443 f. 464.
 579. 636. 875. 935^a. 987. 5015.
 083. 116 f. 133. 257. 266. 269. 354.
 362. 397 f. 402. 430. 454 ff. 645.
 653. 664. 680. 737. 755. 907. 909.
 983. 6001. 015 f. 032. 118. 247.
 307. 311. 318. 320 ff. 333. 401.
 565. 711. 769. 771. 785. 810. 813
 —19. 825 ff. 933 ff. 966 ff. 970.

Abbildung 4: Ausschnitt aus dem Register zu RI XI.

Mit der in der zweiten Zeile des Beispiels genannten Regestennummer „3006“ beginnen die Verweise auf die 3000er-Regestennummern. Der

Nummer folgt aber die Zeichenfolge „025 f.“, wobei man sich die führende Drei hinzudenken muss! Dieser Art ergänzt würde die Reihe der Regestenummern wie folgt aussehen:

2805. 2833 f. 2852. 2855. 2861. 2881 f. 2886. 2919. 2922. 2935. 2938. 3006. 3025 f. 3029. 3031 f. 3035. 3039 ff. 3045. 3047 ff. 3062A. 3103. 3107. 3114. 3118. 3120. 3129 f. 3139 f. 3142. 3144. 3144a. 3147. 3151a. 3160 f. 3223. 3243. 3309. 3441a. 3481 ff. 3495. 3599. 3600 ff. 3688. 3720. 3737 f. 3778 ff. 3782. 3795. 3812. 3814. 3944. 3968. 3973. 4040. 4047. 4049. 4155. 4390 f. 4398. 4417 f. 4443 f. 4464. 4579. 4636. 4875. 4935a. 4987. 5015. 5083. 5116 f. 5133. 5257. 5266. 5269. 5354. 5362. 5397 f. 5402. 5430. 5454 ff. 5645. 5653. 5664. 5680. 5737. 5755. 5907. 5909.

Ähnliche Arbeitsschritte sind bei der Ergänzung der Volltextsuche durch die auf die jeweiligen Regesten verweisenden Lemmata aus dem Register notwendig, um die oben bereits angedeuteten zahlreichen Abkürzungen in den Regestenbänden über die Volltextsuche zu erschließen.

Das heutige Erscheinungsbild der Regesten kann dagegen im wesentlichen erst auf eine ca. 30-jährige Geschichte zurückweisen.

Die bis in die 1970er Jahre allein von der Österreichischen Akademie der Wissenschaften mit bescheidenen personellen und finanziellen Mitteln ausgestatteten *Regesta Imperii* konnten durch die Gründung der Deutschen Regestenkommission (1968) und der bald darauf folgenden Einbindung in das ‚Akademienprogramm‘ von Bund und Ländern zum ersten Mal in der Geschichte des Gesamtunternehmens auf eine solide Basis gestellt werden. Die heute von den Akademien in Mainz, Berlin und Wien betriebenen *Regesta Imperii* verfügen über ein gutes Dutzend Personalstellen und die nötigen Sachmittel. Aufgrund dieser Fundierung konnten in den letzten Jahren brachliegende Forschungsfelder bearbeitet und die Publikationsfrequenz entscheidend erhöht werden. Besonders hingewiesen sei dabei auf die äußerst materialreichen spätmittelalterlichen Abteilungen, denen bei Fehlen einschlägiger Editionen (und auch keiner Aussichten auf solche) ein eigener Darstellungswert beizumessen ist.

933 Februar 27 – 933 Mai	105
<p>933 März, Bergamo 1676</p> <p>Die <i>notarii domnorum regum</i> Andreas und Gumpert sind Zeuge in einer Tauschurkunde der (Kollegiat-)Kirche S. Alessandro in Bergamo, vertreten durch ihren Propst und Custos, den Presbyter Johannes. – Leo <i>notarius</i>. – a. r. Hu. 7, Lo. 2, Ind. 6.</p> <p>Charta (<i>comutacio</i>), ed. Pergamene di Bergamo, Nr. 79 S. 128-129 (Or., vgl. die Taf. 79 im zugehörigen Tafelband); zuvor ed. CD Langob., Nr. 544 Sp. 929-931.</p> <p>Die Datierungsangaben führen einheitlich auf 933 März. – Zur ehemaligen Kathedrale S. Alessandro vor den Stadtmauern in Bergamo vgl. Regg. 1136, 1170 u.ö.</p>	
<p>933 März, villa Villiano (Vigliano) 1677</p> <p>Der Notar Ingelbert schreibt eine Verkaufsurkunde genannter Privatleute. – a. r. Hu. 7, Lo. 2, Ind. 6.</p> <p>Charta (<i>cartula venditionis</i>), ed. GABOTTO, Le più antiche carte di Asti, Nr. 50 S. 87-89.</p> <p>Königsjahr und Indiktion stimmen zu 933 März überein. Bei dem Ausstellort wird es sich um Vigliano d'Asti, Prov. Asti, handeln. – Ingelbert (vgl. auch Reg. 1679) ist noch in den 50er Jahren in Asti bezeugt und hat nie den Titel eines <i>notarius</i> oder <i>iudex domnorum regum</i> geführt; vgl. das Sonderregister. – 933 (?) April wird auch in Ravenna erstmals nach Hugo datiert: CAVARRA, Regesto, S. 440 Nr. 124; vgl. Reg. 1835.</p>	

Abbildung 5: Ausschnitt aus den Karolingerregesten Bd. 3,3 (2006).

<p>1444 Oktober 7, Nürnberg 284</p> <p>Kg. F. belehnt Franz Volckamer (<i>Volkmayr</i>) von Nürnberg, mit seinen Reichslehen, namentlich mit acht Gütlein zu Leinburg (<i>Leymburg</i>), einem Gütlein zu Haimendorf (<i>Hemendorff</i>) und einer Mühle bei Leinburg samt allen Rechten, Nutzen und Zubehör, verleiht ihm daran, was er nach Recht zu verleihen hat, jedoch unbeschadet der Rechte von König und Reich sowie anderer, und bestätigt, daß Franz ihm dafür den gewöhnlichen Lehnseid geleistet hat. <i>An mittichen nach sand Francziscen tag.</i></p> <p>KVr: <i>A.m.d.r. Hermannus Hecht</i>. – KVv: <i>Franz Volkemair</i> (auf der Rückseite des unteren Pressel); <i>Leinpurgk Heymendorff</i> (Empfängervermerk auf der Rückseite).</p> <p>Org. im StA Nürnberg (Sign. Rst. Nürnberg, Heiliggeistspital Nürnberg, Urkunden Nr. 269), Perg., rotes S 11 in wachsfarbener Schüssel mit rotem S 13 rücks. eingedrückt an Ps.</p> <p>Vgl. CHMEL n. 1486 (ohne Monats- und Tagesangabe und mit Ausstellungsort Wien zu 1443 nach RR N fol. 183^o), wonach diese Lehen¹ dem Hans Volckamer verliehen wurden.</p>	
<p>¹ Lediglich abweichend in der Anzahl der Gütlein zu Leinburg, wo von einem Gütlein statt von acht die Rede ist.</p>	

Abbildung 6: Regest aus Heft 14 der *ReggFIII* (insgesamt circa 9 000 Stücke publiziert).

Der gegenwärtige Publikationsstand beläuft sich auf etliche tausend Seiten und ca. 140 000 Regesten.

Die Frage, wie diese beträchtliche, über Jahrhunderte angewachsene Masse in irgendeiner Form ‚einheitlich‘ digital zu präsentieren sei, stand denn auch am Anfang des vor fünf Jahren mit der Bayerischen Staatsbibliothek München begonnenen DFG-Projekts der Retrodigitalisierung der alten und der genuin digitalen Publikation der neueren Bände der *Regesta Imperii*. So waren denn am Anfang die Wünschbarkeiten auch groß: Ausgehend von den Erfahrungen bei der Produktion des ersten elektronisch verfügbaren Regesten-Produkts, der CD-ROM mit den für Heft 1–10 kumulierten Regesten Kaiser Friedrichs III. (1440–1493), sollte der dort verwirklichte Nutzerkomfort analog auf alle anderen Teilprojekte übertragen werden. Doch dies erwies sich, in Anbetracht der schieren Masse und deren Heterogenität und trotz des beträchtlichen finanziellen Einsatzes (OCR-Kosten), schon früh als illusorisch. Um wenigstens das Ziel, alle bisher publizierten Bände im Projektzeitraum ins Netz zu stellen, zu erreichen, wurde die Erschließung, wie man es salopp sagt, flach gehalten. Volltextsuche in teilweise strukturierten Teilen der Regesten (Text, Apparat), Datumssuche und Verlinkung der Literaturzitate mit dem RI-OPAC: Dies war der Kompromiss.

Diese kurze Bestandsaufnahme legt schon nahe, dass mit einiger Sicherheit manches an Informationen aus den gedruckten Bänden bisher digital noch nicht umgesetzt, also verborgen, ist. Manches war sogar, wie wir bei den Vorbereitungen für diese Tagung überrascht feststellen mussten, schon schlichtweg in Vergessenheit geraten. Der Schritt zum Verlorengehen ist da naturgemäß nicht mehr weit. Um was geht es konkret:

Werfen wir zunächst einen Blick auf die immanenten Problemstellen:

Von den 14 Abteilungen der RI werden immer noch 8 (mit wenigen Ausnahmen) allein durch Publikationen der Vorkriegszeit repräsentiert, aus einer Abteilung (IX Wenzel) ist bisher überhaupt noch nichts publiziert. Auch wenn gerade in den nächsten Jahren die Früchte der Neubearbeitungen reichlicher geerntet werden können, so werden wir doch auf unabsehbare Zeit in vielen der älteren Publikationen den zitierfähigen Publikationsstand behalten. Unser besonders Augenmerk wird daher auf der besseren Erschließung dieses Materials liegen müssen.

Albert mag., domherr v. Strassburg
10081.

— Struzius v. Cremona 22 hofrichter
292. 314. 15. 16. 18. 22. 24. 25.
32. 34. 36. 37. 38. 40. 42. 43. 44.
47. 48. 49. 50. 51. 54. 79. 84. 88.
93. 94. 98. 404. 5. 6. 7. 9. 13. 35.
36. 38. 39. 41. 44. 49. 56. 63. 670.
12337. 12441 podesta v. Siena
12315. 16. 39.

— Sudelle 1948.

— v. Summerau 199. 3919. 4089. 90.
92. 96. 98. 4110. 4133. 34. 11028.
14780.

— v. Talmezingen, domherr v. Strass-
burg 8255.

— v. Tanhausen, cämmerer 671. 72.

— v. Than 1071.

— (d. Entartete), landgraf v. Thüringen
u. pfalzgraf v. Sachsen 3463^a. 64.
4885^a. 7790. 8332. 8951. 10181.
11222^a. 11941^a. 48. 77. 12043. 67^a.
14474. 88. 14564^a.

— v. Tyfen 2377.

— (I.) graf v. Tirol 135. 257. 399. 670^a.
89. 93. 716. 21. 22. 1075. 1154.
1603. 1950. 51. 58^a. 2071. 2189.
2279^a. 87. 2359. 2425. 3477. 79.
3854. 4599. 7562. 7876. 8701. 10648.
10784. 10860. 74^a. 79. 11030. 11121.
11309. 13. 59. 11649. 11925. 12280.
12308. 13056^a. 91^a. 13204. 9. 14004.
14639 (?). 14728^a. 15091.

— de la Tour, truchsess v. Arelat u.
Vienne 5489. 5503.

— de Treffoso 393.

— bisch. v. Treviso 9163. 14047. 14111.
11^a.

— bisch. v. Trient 1075. 80. 1154. 73.
79. 82. 93. 1208. 9. 15. 17. 18. 20.
22. 23. 29. 30. 32. 33. 35. 36. 37.
39. 40. 41. 43. 47. 52. 59. 61. 66.
79. 92. 94. 97. 1301. 6. 7. 8. 9. 12.
18 N. 60. 1423. 25. 26. 33. 35. 37
—41. 47. 49. 50. 51. 57—63. 71. 74.
1544. 1939. 3871. 10860. 64^a. 79.
12842. 14670. 15053 reichslegat in
Tuscien 6541. 12861^a. 63.

— v. Trimberg 4416. 4568. 4867. 11415.

— de Turixella, podesta v. Vercelli
14045.

— markgraf v. Tuscien 358.

— Ubertini 1244.

— v. Ulm, mag., minorit 4407.

— v. Utinheim 718.

— bisch. v. Vercelli 12242.

— v. Vipeche, probst v. Magdeburg
4373.

— de Voorne 4989.

— v. Vroburg 131. 32. 45.

— v. Wangen 135.

— graf v. Weichselberg 1956. 11677.

— v. Werbinwac, ritter 3845.

— abt v. Werden 5304.

— (III.) graf v. Wermigerode 35. 101.
3. 9. 31. 63. 64. 14619.

— graf v. Wiehe 4191. 4250 graf v.
Rabenswald 11571.

— truchsess v. Wirzburg 4071.

— v. Wolkenberg 4780.

— v. Wunsiedel 3912.

Abb. 7: Ausschnitt aus dem Personennamenregister.

Folgende Informationseinheiten seien als signifikant hervorgehoben:

- Register,
- Verzeichnisse,
- Konkordanzen,
- Literaturverlinkung,
- Vorworte/Einleitungen,
- Nachträge (mit Mehrfachnumerierungen).

Die in den älteren Bänden nur kümmerlich entwickelte Textstruktur lässt den beigegebenen Apparaten (Verzeichnissen) einen hohen Stellenwert zukommen. Am Beispiel der schon vorhin gezeigten Auszüge aus den Stauferregesten lässt sich dies gut veranschaulichen.

Das für diese Abteilung, wie längst nicht für alle, vorliegende Register enthält in beträchtlicher Zahl Metainformationen zu den Regesten (Orts und Personenidentifikationen), liegt aber selbst wiederum in einem für die digitale Aufbereitung problematischen Format vor: Kürzungen in den textlichen Teilen und, weit unangenehmer, Kürzungen in den Regestennummern. Allein um hier zu einer ‚einfachen‘ Verlinkung von Regestennummern und Text-Register zu gelangen, bedarf es erheblichen Aufwandes. Hier wird es in absehbarer Zeit Übergangslösungen geben, derart, dass die Register mit dem Volltext gemeinsam durchsuchbar sein werden.

Eine Vielzahl von Informationen ist auch in den mannigfaltigen den älteren Bänden beigegebenen Verzeichnissen und Konkordanzen verborgen. Besonders augenscheinlich ist das bei den Referenzierungen auf ältere Bearbeitungen der Regesta-Ausgaben oder ersatzweise herangezogene Werke (zum Beispiel Chmels *Regesta Friderici* für die Regesten Friedrichs III.).

Problematisch, gleichsam nur als Informationsbrocken, kommen die Quellenbelegstellen und die Literaturangaben in den älteren Bänden daher. Hier gibt es außer gelegentlich summarischen Übersichten in den Einleitungen keinerlei auflösende, die vollständigen bibliographischen Angaben enthaltende Verzeichnisse. Aus diesem Grunde haben wir in den *RI-Online* auch diesem Bereich zunächst die meiste Aufmerksamkeit gewidmet, ist hier doch gegenüber den gedruckten Bänden ein deutlicher Mehrwert zu erzielen. Dreh- und Angelpunkt ist dabei der *RI-OPAC*, der aufgrund seiner Größe und thematischen Bandbreite längst ein Eigenleben bezüglich der Literaturversorgung aller an mediävistischen Themen Interessierter führt.

Dass diese mittlerweile weltweit umfangreichste Fachbibliographie einen hohen Imagewert für das Gesamtunternehmen besitzt, ist ein willkommener Nebeneffekt. Für die Binnenerschließung der Regesten dient der *RI-OPAC* insoweit, dass jedes zitierte Werk in ihm enthalten ist und über einen eindeutigen, unveränderlichen Kurztitel die Verlinkung gewährleistet. Für die Regestenbearbeiter besitzt der *RI-OPAC* natürlich wesentliche Synergieeffekte: Es gibt eine zentral geführte Datei, eindeutige bibliographische Angaben, Neueingaben können künftig direkt im Netz vorgenommen werden. Mit dem *RI-OPAC* durchbrechen wir aber auch schon weitgehend das Problem des ‚Vergessens‘, denn in ihm sind bereits große Teile der Arbeitsbibliographien sämtlicher Arbeitsstellen enthalten!

Lassen sich die inneren verborgenen Schätze bei dauerhaftem personellen und sachlichen Engagement im Laufe der Zeit heben, so gibt es einen äußeren Bereich von einschlägigen Informationen, in dem die Chancen schlechter stehen. Dass wir diesen Bereich überhaupt künftig möglicherweise sinnvoll einbeziehen können, verdanken wir allein den digitalen Möglichkeiten. Gemeint sind hier Informationen, die sich in welcher Hinsicht auch immer auf die Regesten beziehen, seien es Corrigenda, Zusätze im Text, im Apparat, in den Literaturangaben, in der kritischen Bewertung der Historiographie, in der diplomatischen Qualität der Urkunden (Fälschungen), Informationen, die an ganz unterschiedlichen Stellen (ich betone: noch) vorhanden sein können. Zunächst ist dabei natürlich an die Urheber der Regesten, die Autoren und Herausgeber, zu denken, die – sofern sie ihrem Sujet verbunden geblieben sind respektive bleiben – mit einer gewissen Systematik dieses ergänzende Material sammeln. Ferner kann man an Forscher denken, die aus eigenem Forschungsinteresse das Regesten-Material als Grundlage genommen und im Laufe der Zeit weiterverarbeitet haben (das können durchschossene Exemplare wie das in unseren Fachkreisen bekannte Exemplar der *RI VIII* des Giessener Mediävisten Peter Moraw sein, natürlich auch Karteien oder Dateien an den verschiedensten Orten). Des Weiteren ist an Korrespondenzen zwischen Bearbeitern und Rezipienten, Gutachtern und so weiter zu denken und schließlich an die nun digitalen Möglichkeiten, Verbesserungen und Kritik online anzubringen.

Angesichts der fast zwei Jahrhunderte umfassenden Publikationszeit muss man realistischerweise von etlichen Informationsverlusten ausgehen, denn: Weder gab es bis in die neuere Zeit eine nennenswerte Archivierung von Arbeitsmaterialien, noch Mittel und Möglichkeiten, etwa zufällig anfal-

lendes Material zu publizieren. Als besonders nachteilig wirkt sich hier die Organisationsstruktur der *Regesta Imperii* aus: im Gegensatz etwa zur MGH handelt es sich bei den RI nicht um ein auf Dauerbetrieb angelegtes Institut, sondern um einen Verein, dessen Projekte auf Endlichkeit angelegt sind und die von einer einzigen Personalstelle aus koordiniert werden. Die Wahrnehmung in der Öffentlichkeit ist naturgemäß eine andere und so ist es kein Zufall, dass auch die *RI* (respektive ihre Geschäftsstelle und Einzelunternehmungen) Materialien der oben angegebenen Art erhalten und so gut es geht sammeln. Als ein besonders gelungenes Beispiel für die Rettung und Nutzung von Vorarbeiten sei das Projekt der Regesten Kaiser Ludwigs des Bayern erwähnt, dass sich auf den im Münchner Hauptstaatsarchiv verwahrten Nachlass von Friedrich Bock stützen kann.

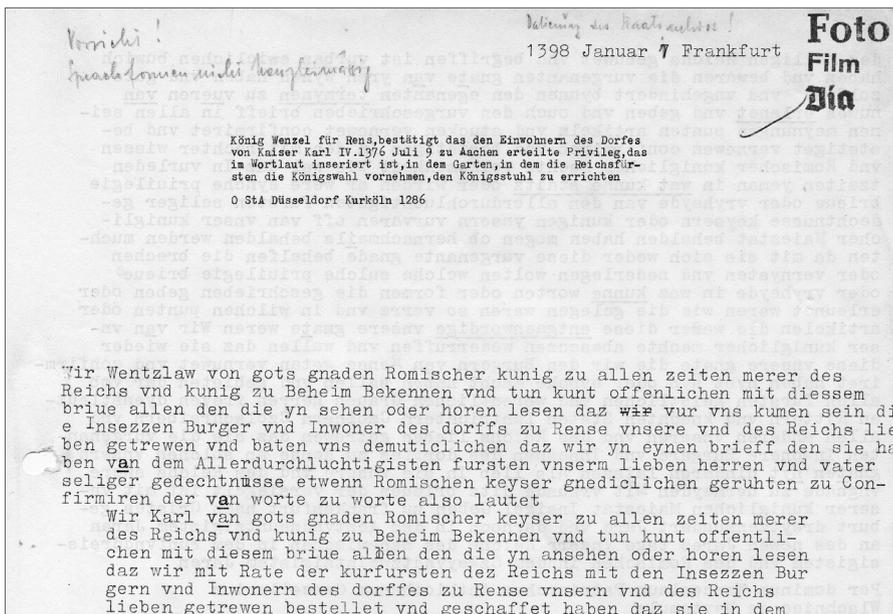


Abbildung 8: Transkription aus dem Nachlass Hanisch.

Ein weiterer interessanter Nachlass wird seit 20 Jahren in der Mainzer Geschäftsstelle der Regestenkommission verwahrt: die Sammlung Hanisch. Hierbei handelt es sich um die von dem ehemaligen Veichtaer Universitäts-

bibliothekar Wilhelm Hanisch privat und auf eigene Kosten betriebenen, durchaus auf eine Regestenpublikation abzielende Sammlung von urkundlichen Belegen zu Wenzel von Böhmen (1379–1400/10). Ohne auf die verwickelte Geschichte dieser Sammlung, zu der man auch diejenige einer ähnlichen, von Ivan Hlaváček in Prag aufgebauten, zählen müsste, im Einzelnen eingehen zu wollen: Wilhelm Hanisch hat im Laufe von 40 Jahren akribisch die europäischen Archive nach Wenzeliana abgesucht und in Wort und Bild circa 5 000 Stücke eruiert.

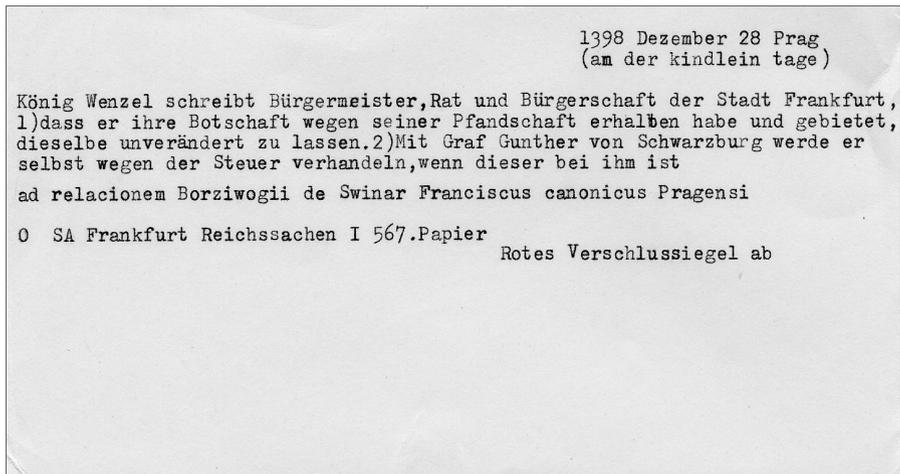


Abbildung 9: Karteikarte aus dem Zettelkasten der Sammlung Hanisch.

Wie bereits angedeutet, sind bisher noch keine Regesten Wenzels erschienen, die von der Österreichischen Akademie federführend betriebene Abteilung wird in nächster Zeit zwar mit der Publikation beginnen, jedoch aus verschiedenen Erwägungen heraus analog den Ludwig- und Friedrich-Regesten nach dem Provenienzprinzip. Noch auf lange Zeit wird also die Hanisch-Sammlung wertvoll bleiben. Daher planen wir nun im Rahmen der *RI-Online* eine digitale Publikation des Materials.

Es handelt sich in erster Linie um eine auf der Grundlage der Arbeitskarteien erstellten chronologischen Datei der Urkunden Wenzels. Diese wurde Ende der 1980er Jahre in einem heute obskur anmutenden Programm namens ‚AskSam‘ erfasst, dass nun wiederum migriert werden

- Verteiltes Arbeiten ist möglich,
- für die Bearbeitung des Corpus sind keine speziellen EDV- oder HTML-Kenntnisse nötig,
- die Publikation wächst ständig, alle Änderungen und Ergänzungen sind sofort im Netz sichtbar.

Einige dieser Punkte machen Typo3 auch für die Ausgestaltung der Homepage der Regestenkommission interessant. Insbesondere die Möglichkeit, einzelne Abschnitte des Seitenbaumes nur bestimmten Nutzergruppen zur Bearbeitung zu überlassen, ermöglichen es interessierten Projekten, selbst über die Inhalte der sie betreffenden Seiten zu entscheiden, ja diese auch selbst zu gestalten. Dies führt im Idealfall nicht nur zu umfassenderen, sondern auch zu einer aktuelleren Seitengestaltung. Neben einer gegebenenfalls notwendigen Schulung der Mitarbeiter steht der Seitenneugestaltung ansonsten nichts im Wege.

Mit diesem Ausblick auf die aktuellen Schwerpunkte unseres Digitalisierungsvorhabens möchten wir schließen, nicht ohne die Hoffnung, Ihnen einen erhellenden Einblick in das ‚dunkle‘ Mittelalter respektive in die Erforschung der Grundlagen desselben gegeben zu haben.

Virtuelle Zusammenführung und inhaltlich-statistische Analyse der überlieferten Reichskammergerichtsprozesse

Bernd Schildt

Die archivalische Hinterlassenschaft des Reichskammergerichts ist gekennzeichnet durch eine überaus disparate Quellenüberlieferung. So finden sich Verfahrensakten des Reichskammergerichts nicht nur in zahlreichen deutschen, sondern auch in einer ganzen Reihe ausländischer Archive, so zum Beispiel in Frankreich,¹ Polen,² den Niederlanden,³ Belgien⁴ und Österreich.⁵

In den Jahren 1847–1852 wurden circa 70 000 Prozessakten des 1806 ohne ein vergleichbares Nachfolgegericht untergegangenen Reichskammergerichts unter die Mitglieder des Deutschen Bundes verteilt. Diese Zersplitterung wurde nach 1852 in einer Reihe von Bundesgliedern durch weitere Aufteilungen noch fortgesetzt.⁶ An eine systematische Erforschung der Tätigkeit des Reichskammergerichts war infolge dieser Aufteilung der Prozessakten auf mehr als 40 in- und ausländische Archive lange Zeit kaum zu denken. Mittlerweile ist die Neuverzeichnung der Prozessakten nach den so genannten DFG-Richtlinien⁷ in den meisten deutschen Verwahrarchiven so

¹ Im Departementalarchiv Straßburg für die Landgrafschaft Elsaß, das Herzogtum Lothringen und weitere angrenzende ehemalige Reichsstände, im Departementalarchiv Metz für das Herzogtum Lothringen sowie im Stadtarchiv Straßburg für die Reichsstadt Straßburg und im Departementalarchiv Besancon für Stadt und Erzbistum Besancon.

² Im Wojewodschaftsarchiv Stettin für das Herzogtum Pommern.

³ Im Reichsarchiv Limburg (Maastricht) für das Königreich der Niederlande.

⁴ Im Staatsarchiv Lüttich für das Königreich Belgien.

⁵ Im Haus-, Hof- und Staatsarchiv Wien für das Erzbistum Salzburg sowie die Fürstbistümer Trient und Brixen.

⁶ Vgl. Walter Latzke: Das Archiv des Reichskammergerichts. In: Zeitschrift der Savigny-Stiftung für Rechtsgeschichte. Germanistische Abteilung 78 (1961). S. 321–326.

⁷ Die *Grundsätze für die Verzeichnung von Reichskammergerichtsakten* sind abgedruckt in: Martin Ewald: Inventarisierung von norddeutschen Beständen des Reichskammergerichts. In: Der Ar-

weit vorangeschritten, dass eine wesentliche Verbesserung der Forschungslage eingetreten ist.⁸

Das Projekt zur Erfassung des überlieferten Prozessmaterials des Reichskammergerichts in einer Datenbank ist angesiedelt an der Schnittstelle archivalischer Aufbereitung und Sicherung historischer Massendaten – der Wissensbestände – und deren systematischer inhaltlicher Erschließung und Analyse auf der Basis der in den meisten Archiven bereits vorab geleisteten Verzeichnungsarbeiten.

Ziel des Reichskammergerichts-Datenbankprojektes (*RKG-Datenbank*) ist die virtuelle Zusammenführung und erste inhaltliche Analyse der überlieferten Prozessakten des Reichskammergerichts. Ausgangspunkt ist eine weitgehend homogene archivalische Erschließung des Quellenmaterials. Seit 1981 sind nunmehr 25 *Inventare der Akten des Reichskammergerichts* gemäß den Vorgaben der genannten DFG-Richtlinien erstellt worden.⁹ Diese Inventare sind zwar weitgehend nach einheitlichen Grundsätzen erarbeitet, stehen aber dennoch letztlich isoliert nebeneinander.¹⁰ Durch die zusätzliche, methodisch neuartige Erfassung in einer relationalen Datenbank werden deutlich verbesserte Voraussetzungen für eine umfassende inhaltliche und räumlich wie zeitlich übergreifende Untersuchung aller in den Archivinventaren verzeichneten Reichskammergerichtsprozesse geschaffen.

Der Wert der Datenbank ist inhaltlich natürlich abhängig von der Qualität der Inventare. Gleichwohl geht sie aber nicht nur durch ihre virtuelle Verknüpfung, sondern – wie noch zu zeigen sein wird – auch sachlich deutlich über deren Informationsgehalt hinaus. Unter Einbeziehung der bereits in einer Access-Datenbank erfassten Urteilsbücher (AR 1)¹¹ wurde entsprechend den inhaltlichen Möglichkeiten, wie sie sich aus der Verzeich-

chivar 33 (1980), Sp. 481–482, Anhang.

⁸ Vgl. Jürgen Weitzel: Das Inventar der Akten des Reichskammergerichts (Forschungsbericht). In: Zeitschrift für Neuere Rechtsgeschichte (ZNR) 21 (1999). S. 408–416. – Zuletzt Bernd Schildt: Inhaltliche Erschließung und ideelle Zusammenführung der Prozessakten des Reichskammergerichts mittels einer computergestützten Datenbank. In: ZNR 25 (2003). S. 269–290.

⁹ Davon sind 24 Inventare abgeschlossen. Für die Bestände in München sind bislang zwölf Bände erschienen.

¹⁰ So wurden vier weitere, ältere Verzeichnungen mit insgesamt rund 10 000 Prozessen zwar nachträglich in diese Reihe aufgenommen, folgen in wesentlichen Punkten aber nicht den DFG-Richtlinien (Nr. 1 für die Regierungsbezirke Koblenz und Trier der preußischen Rheinprovinz, Landeshauptarchiv Koblenz; Nr. 2 für die Preußische Provinz Westfalen, Staatsarchiv Münster; Nr. 3 für das Herzogtum Braunschweig, Staatsarchiv Wolfenbüttel; Nr. 4 für den Regierungsbezirk Stade der preußischen Provinz Hannover, Staatsarchiv Stade). Vgl. auch Weitzel, wie Anm. 8, *passim*.

nungstiefe der Archivinventare ergibt, und unter Nutzung der Erfahrungen älterer Versuche einer datenbankmäßigen Erschließung¹² von Prozessakten des Reichskammergerichts – gemäß den Möglichkeiten moderner Datenverarbeitung – ein mittlerweile ausgereiftes Konzept entwickelt.

Die hier kurz beschriebene Datenbank¹³ soll keineswegs dazu dienen, die gedruckten Inventare der Akten des Reichskammergerichts überflüssig zu machen; vielmehr ist daran gedacht, dass beide Formen der Aufbereitung des Quellenmaterials einander ergänzen. Für Einzelfallprobleme, die sich einer formalisierenden Erfassung in der Datenbank entziehen (zum Beispiel Details der Sachverhaltsbeschreibungen und Beweismittel, sprachliche Besonderheiten bei den Namen der Parteien oder Berufsbezeichnungen und ähnlichem), bleibt der Rückgriff auf die Inventare nach wie vor erforderlich und sinnvoll. Gegebenenfalls wäre daran zu denken, die Datenbank künftig mit bereits vorhandenen oder noch zu erstellenden¹⁴ digitalisierten Volltextversionen der Inventare der Akten des Reichskammergerichts zu verlinken. Dies könnte selbstverständlich nur in einem komplexen Zusammenwirken mit den beteiligten Archiven erfolgen.

Bei den Inventarisierungsarbeiten wurde mit Blick auf die nicht mehr als Aktenbestände vorhandenen Reichskammergerichtsverfahren in den einzelnen Archiven unterschiedlich verfahren. Teilweise wurden diese Prozesse mit dem Vermerk ihres Verlustes versehen aufgenommen,¹⁵ teilweise wurde auf deren Erfassung vollständig verzichtet, so beispielsweise in Darmstadt, wo circa 90 Prozent des Aktenbestandes bereits im 19. Jahrhundert kassiert worden sind.¹⁶ Vor diesem Hintergrund werden auch die nur

¹¹ Hans Schenk: Reichskammergericht Bestand AR 1 Urteilsbücher, Datenbank und Begleitheft (Findbücher zu den Beständen des Bundesarchivs 52). Koblenz 1995.

¹² Vgl. Filippo Ranieri: Recht und Gesellschaft im Zeitalter der Rezeption. Eine rechts- und sozialgeschichtliche Analyse der Tätigkeit des Reichskammergerichts im 16. Jahrhundert (Quellen und Forschungen zur Höchsten Gerichtsbarkeit im Alten Reich 17 I/II). Köln, Wien 1985.

¹³ Vgl. Im Einzelnen Schildt, wie Anm. 8, passim. Zum jeweils aktuellen Verzeichnungsstand vgl. Adresse: <http://www.ruhr-uni-bochum.de/rkg-forschung/>.

¹⁴ Das dürfte für eine ganze Reihe von Inventaren mit einer beträchtlichen Zahl von Prozessen erforderlich sein.

¹⁵ Im Bremer und Frankfurter Inventar werden diese Verfahren zum Beispiel kursiv gesetzt, in München sind sie mit dem Hinweis „Akt makul.“ versehen.

¹⁶ Andrea Korte-Böger und Cornelia Rösner-Hausmann: Reichskammergerichtsakten im Hessischen Staatsarchiv Darmstadt und im Gräflich Solmsischen Archiv in Laubach (Repertorien des Hessischen Staatsarchivs Darmstadt 31, Inventar der Akten des Reichskammergerichts 15). Darmstadt 1990. S. IX.

nach dem *Wetzlarer Generalrepertorium* oder anderen Findhilfsmitteln erschließbaren Verfahren, die aus unterschiedlichsten Gründen aktenmäßig nicht mehr überliefert sind, in der Datenbank erfasst. Soweit diese Prozesse in den Inventaren nicht oder nur unvollständig verzeichnet wurden, müssen die erforderlichen Informationen anhand der entsprechenden Spezialrepertorien ermittelt werden. Insoweit geht die *RKG-Datenbank* auch quantitativ über die Archivinventare hinaus. Die auf diese Weise erfassbaren Angaben sind zwar gegenüber den nach DFG-Richtlinien vollständig inventarisierten Aktenbeständen inhaltlich verkürzt, können aber separat abgefragt werden. Damit lassen sich inhaltliche Fehlinterpretationen statistischer Art vermeiden.

Gemäß den DFG-Richtlinien sollen den einzelnen Inventaren umfangreiche Indizes beigelegt werden. Es handelt sich dabei um einen kombinierten Personen- und Ortsindex, um einen Sachindex, ein Verzeichnis der Prokuratoren, ein chronologisches Verzeichnis der Prozesse und ein Verzeichnis der Vorinstanzen und Juristenfakultäten. Diese für die Benutzung der Inventarbände überaus wichtigen Indizes weisen in ihrer Qualität indes deutliche Unterschiede auf.¹⁷ Fraglich ist ferner, ob es jemals – wie vorgesehen – zur Erstellung eines Gesamtindex kommen wird. Da dies aber auf jeden Fall erst nach Abschluß der Inventarisierungsarbeiten erfolgen kann, wird es noch geraume Zeit an übergreifenden, die einzelnen Archivinventare im Zusammenhang erschließenden Indizes fehlen.¹⁸

Durch die *RKG-Datenbank* werden bis auf den Sachindex aber *alle* Indizes für die bislang erfassten Bestände bereits jetzt zur Verfügung gestellt. So werden Personen und Orte durch die Abfragen ‚Name der Parteien‘ bzw. ‚Sitz der Parteien‘, das Verzeichnis der Prokuratoren durch ‚Name des Vertreters‘ und das chronologische Verzeichnis der Prozesse durch ‚Beginn des Verfahrens‘ bzw. ‚Ende des Verfahrens‘ erschlossen. Das Verzeichnis der Vorinstanzen und Juristenfakultäten kann in den Feldern ‚Name des Gerichts‘ bzw. ‚Beweismittel‘ und ‚Beweismittel/Erläuterung‘ jeweils separat abgefragt werden.¹⁹

¹⁷ Vgl. Weitzel, wie Anm. 8, S. 410.

¹⁸ Für den größten Bestand an Reichskammergerichtsprozessakten im Hauptstaatsarchiv München (circa 14 000 Verfahren) ist nach gegenwärtigem Stand mit einer vollständigen Publizierung des Inventars erst in etwa 15 bis 20 Jahren zu rechnen.

¹⁹ Vgl. im Einzelnen Anlage 4: Suchmaske.

Die spätere Einfügung oder auch nur Verlinkung eines noch zu erarbeitenden umfassenden Sachindexes ist bei entsprechender Absprache und Koordinierung mit den Archiven unproblematisch möglich. Denkbar wäre auch eine sukzessive Einarbeitung anhand der vorhandenen Indizes; dies bedürfte aber wohl noch grundsätzlicher Überlegungen, vor allem unter dem Gesichtspunkt von Aufwand und Nutzen.

Die Sachverhaltsbeschreibungen sowie die Angaben zu den Beweismitteln in den Volltextversionen der Inventare lassen sich mit den in der Datenbank verwendeten Kategorien strukturell nicht vergleichen. Während bei den Inventaren eher der Einzelfall in den Blick genommen wird, ist die Datenbankeinfassung vornehmlich der Herstellung der Vergleichbarkeit möglichst vieler – im Idealfall aller – Einzelfälle verpflichtet. Daraus ergeben sich zwangsläufig Konsequenzen für die Nutzungsmöglichkeiten.

Die vom Wortlaut der Texte ausgehenden Indizes und auch der vorgesehene Generalindex ermöglichen lediglich den Zugriff auf einzelne Verfahren anhand bestimmter Begrifflichkeiten. Auch die in digitalisierter Form vorliegenden Inventare lassen nur eine Suche anhand des dort jeweils verwendeten Vokabulars zu. Demgegenüber besteht der Vorteil der Datenbank darin, dass sie den unterschiedlichen Sprachgebrauch in den Inventaren, und zwar in allen, insofern vereinheitlicht, als sie vergleichbare, überwiegend, aber nicht nur, rechtshistorische Sachverhalte und Termini in formalisierter Form erfasst und dadurch auch erst sinnvoll abfragefähig macht.²⁰

So werden aus den Sachverhaltsbeschreibungen der Archivinventare und den Darin-Vermerken nach einheitlichen, formalisierten Kriterien detaillierte Angaben zum juristischen Inhalt der Prozesse, zu den Beweismitteln, den verschiedenen Prozessformen und ähnlichem erfasst. Dies ermöglicht eine statistisch relevante, sachliche Erschließung unter allgemeinen und speziell rechtshistorischen Gesichtspunkten und dann – in einem zweiten Schritt – den zielsicheren Zugriff auf spezifische Einzelprobleme. Es ergeben sich damit ganz neue Möglichkeiten für die Erschließung des überlieferten Quellenmaterials unter inhalts- bzw. sachbezogenen Kriterien. Vergleichbares, das in den Inventaren unterschiedlich umschrieben wird, kann einheitlich erfasst und folglich vom Nutzer später auch abgefragt

²⁰ Vgl. zu dieser Problematik auch Gert Polster: Die elektronische Erfassung des Wolfschen Repertoriums zu den Prozessakten des Reichshofrates im Wiener Haus-, Hof- und Staatsarchiv. In: Mitteilungen des Österreichischen Staatsarchivs 51 (2004). S. 635–649. S. 644.

werden.²¹ Angesichts dieses – was die archivalische Quellenaufbereitung anbelangt – insgesamt erfreulichen Befundes, besteht inzwischen die Möglichkeit, diese Quellen nunmehr auch als Grundlage für konkrete Forschungen über den Einzelfall oder Fallstudien hinaus zu nutzen.

Ein weiterer Vorteil der Datenbank besteht darin, dass sie die Voraussetzungen von quantitativen Analysen ganz erheblich verbessert, da sie die Recherche unter verschiedenen Gesichtspunkten gleichzeitig ermöglicht. Die Datenbank gestattet die Abfrage nach 38 Kriterien, die frei miteinander kombiniert werden können.²² Damit wird ein zielgenauer Zugriff nicht nur unter sachlichen, sondern auch unter personellen, räumlichen, strukturellen (zum Beispiel Gerichte, Vorinstanzen) und zeitlichen Gesichtspunkten möglich.

Neben diesen komplexen Abfragemöglichkeiten anhand der 38 Auswahlkriterien ist die Datenbank – entsprechend den Bedürfnissen spezieller Forschungsinteressen – prinzipiell auch offen für weitergehende Abfragen. Für die nicht formalisierbaren Individualeingaben erfolgt dies nach dem bekannten Prinzip der ‚Fernglassuche‘, innerhalb der selbstergänzenden Kombinationsfelder kann mittels der *Sternchensuche* recherchiert werden. Über diese Einzelabfragen hinaus ist auch die Schaffung neuer komplexer Abfragen grundsätzlich denkbar.

Prinzipiell hängt der statistische Wert von Datenbankabfragen selbstverständlich maßgeblich von der Anzahl der eingearbeiteten Prozesse ab. Die Angaben zu abgeschlossenen Inventaren haben aber, bezogen auf deren Zuständigkeitsbereich, eine eigenständige Bedeutung. Sie reicht teilweise auch über den Zuständigkeitsbereich einzelner Inventare hinaus.²³ Die Form der Erfassung in einer Datenbank macht es also möglich, bereits auf der Grundlage einzelner Inventare, deren Erfassung abgeschlossen ist, spezielle Untersuchungen durchzuführen, da es sich bei diesen Daten um die geschlossene Überlieferung eines exakt bestimmten politisch-territorialen

²¹ Damit geht die *RKG-Datenbank* deutlich über die digitale Fassung des Wolf'schen Repertorioms (vgl. Anm. 20) hinaus.

²² Vgl. im Einzelnen Anlage 4: Suchmaske.

²³ So wird insbesondere mit der vollständigen Erfassung des württembergischen und bayerischen Bestandes eine erhebliche Vermehrung der Prozesse reichsstädtischer Provenienz verbunden sein. Vgl. dazu die Übersicht bei Kathrin Dirr und Torsten Joecker: Die Inanspruchnahme des Reichskammergerichts durch die Reichsstädte. In: *Prozesspraxis im Alten Reich. Annäherungen – Fallstudien – Statistiken*. Hg. von Anette Baumann und anderen (Quellen und Forschungen zur Höchsten Gerichtsbarkeit im Alten Reich 50). Köln, Weimar, Wien 2005. S. 119–136.

bzw. räumlich-administrativen Bereiches handelt. Insoweit haben diese Teilergebnisse des Datenbankprojekts durchaus einen unmittelbaren Eigenwert für entsprechende Forschungsvorhaben.²⁴

Dass die *RKG-Datenbank* kein Selbstzweck ist, lässt sich anhand erster Ergebnisse mittlerweile verifizieren. Neben der in hohem Maße ‚nur‘ forschungsvorbereitenden Funktion des Datenbankprojekts wurden und werden, verbunden mit intensiven archivalischen Forschungen, zunehmend auch inhaltliche Ergebnisse erbracht. Das hängt eng mit dem personellen Konzept des Gesamtprojekts zusammen.

Die sehr personalintensiven Arbeiten bei der Eingabe der Daten in die Datenbank werden von speziell geschultem Personal – Rechtsreferendaren bzw. Volljuristen – geleistet. Deren Motivation wird durch die Möglichkeit zur Promotion entscheidend gefördert. Selbstverständlich stehen die jeweiligen Dissertationsthemen regelmäßig in engem Zusammenhang mit dem vom betreffenden Mitarbeiter bearbeiteten Inventar. Grundsätzlich erfolgt bei diesen Arbeiten neben einer quantitativen Analyse des bearbeiteten Inventars anhand der Datenbank zwingend die Bearbeitung eines weithin frei wählbaren rechtshistorischen Sachthemas, das auf der Basis archivalischer Quellenstudien zu bearbeiten ist. Auf diese Weise sind inzwischen unmittelbar aus dem Projekt drei Dissertationen hervorgegangen.

Mittlerweile veröffentlicht sind die beiden Arbeiten von Anna-Maria Savelsberg zur Pfändungskonstitution der Reichskammergerichtsordnung von 1555²⁵ und von Kathrin Dirr-Jansen zu hoheitsrechtlichen Streitigkeiten der Kölner Erzbischöfe mit der Stadt Köln.²⁶ Quellengrundlage sind die überlieferten Verfahrensakten des Reichskammergerichts im Staatsarchiv Darmstadt bzw. im Stadtarchiv Köln. Im Erscheinen begriffen ist die Arbeit von

²⁴ Welche Möglichkeiten derartige Teilergebnisse für die Untersuchung spezieller Fragestellungen bieten, zeigen die bereits vollständig in der Datenbank erfassten Inventare. Vgl. zum Beispiel Kathrin Dirr: Die Bedeutung limitierter Reichskammergerichtsprivilegien der Reichsstadt Köln für die Entwicklung der städtischen Gerichtsverfassung. In: *Zeitenblicke*, 3 (2004). Nr. 3 vom 13.12.2004, Adresse: http://www.zeitenblicke.de/2004/03/schildt/index.html#headline_quote_letzte_Einsichtnahme_am_04.05.2006. – Torsten Joecker: Reichsstädte als Sitz des Reichskammergerichts. In: Ebd. – Bernd Schildt: Datenbank Reichskammergerichtsakten. In: Ebd.

²⁵ Anna-Maria Savelsberg: Die Pfändungskonstitution der RKGO 1555, Teil 2 Tit. 22 als ein landesherrliches Mittel zum Ausbau der Territorialstaatlichkeit. München 2004.

²⁶ Kathrin Dirr: Hoheitsrechtliche Streitigkeiten zwischen den Kölner Erzbischöfen und der Stadt Köln auf Grundlage reichskammergerichtlicher Verfahren des 16. und 17. Jahrhunderts (Rechtshistorische Reihe 313). Frankfurt am Main 2005.

Christian Vajen zur Problematik der rechtlichen Anerkennung reformierter Reichsstände vor dem Westfälischen Frieden.²⁷

Neben den genannten Dissertationen – weitere sind in Vorbereitung und stehen zum Teil kurz vor dem Abschluss – kann in diesem Zusammenhang auch auf die monographische Abhandlung des Autors zur Entwicklung der Zuständigkeit des Reichskammergerichts (2006)²⁸ und den Aufsatz von Kathrin Dirr und Torsten Joecker zur Inanspruchnahme des Reichskammergerichts durch die Reichsstädte²⁹ verwiesen werden.

Naturgemäß sind Datenbanken sehr viel flexibler handhabbar als gedruckte Bestandsverzeichnisse oder Inventare. Die Aufnahme neuerer Forschungsergebnisse ist jederzeit und unproblematisch möglich. Dem trägt auch die *RKG-Datenbank* durch eine besondere Rubrik – ‚Verweise‘³⁰ – Rechnung, in der – soweit ersichtlich – Querverbindungen zwischen den einzelnen Verfahren kenntlich gemacht werden, auf verwandtschaftliche Beziehungen zwischen den Beteiligten hingewiesen wird und vor allem Literaturangaben zu einzelnen Prozessen eingearbeitet werden. Insbesondere die in den Archivinventaren nachgewiesene Literatur ist bereits komplett verarbeitet worden. In einer späteren Bearbeitungsphase ist vorgesehen, systematisch nach einschlägiger – auch zeitgenössischer – Literatur zu den einzelnen Fällen zu suchen (zum Beispiel in den *Wetzlarer Nebenstunden*³¹). Ferner wird zurzeit daran gearbeitet, die jüngst erschienenen gedruckten Relationen und Voten des Reichskammergerichts vom 16. bis 18. Jahrhundert in die Datenbank zu integrieren.³² Selbstverständlich kann auch auf

²⁷ Christian Vajen: Die rechtliche Anerkennung reformierter Reichsstände durch den Religionsfrieden vor Abschluß des Westfälischen Friedensvertrages – Eine Darstellung auf der Basis lippischer Reichskammergerichtsprozesse (Bochumer Forschungen zur Rechtsgeschichte 1). Aachen 2006.

²⁸ Bernd Schildt: Die Entwicklung der Zuständigkeit des Reichskammergerichts. Von der Kayserlichen Cammer=Gerichts=Ordnung Anno 1495 zum Concept der Cammer=Gerichts=Ordnung vom Jahr 1613 (Schriftenreihe der Gesellschaft für Reichskammergerichtsforschung 32). Wetzlar 2006.

²⁹ Vgl. Anm. 23.

³⁰ Vgl. im Einzelnen Anlage 4: Suchmaske.

³¹ *Wetzlarische Nebenstunden. Worinnen auserlesene beyrn höchstpreiflichen Cammergericht entschiedene Rechts-Händel zur Erweiter- und Erläuterung der teutschen in Gerichten üblichen Rechts-Gelehrsamkeit angewendet werden.* Hg. von Johann Ulrich von Cramer, 128 Teile (1755–1773).

³² Anette Baumann: Gedruckte Relationen und Voten des Reichskammergerichts vom 16. bis 18. Jahrhundert. Ein Findbuch (Quellen und Forschungen zur Höchsten Gerichtsbarkeit im Alten Reich 48). Köln, Weimar, Wien 2004.

Zusammenhänge zu Parteiakten oder anderen Quellen hingewiesen werden. Insoweit ist es möglich, sukzessive neuere Forschungsergebnisse regelmäßig in der Datenbank ‚nachzutragen‘.

Die *RKG-Datenbank* ist prinzipiell offen für die Einbeziehung der Prozesse des anderen obersten Reichsgerichts – des Reichshofrats (circa 70 000 Verfahren). Allerdings sind die archivalischen Vorarbeiten für den Reichshofrat deutlich schlechter als für das Reichskammergericht, da bislang nur für einen Bruchteil der Prozesse – im Wesentlichen etwa die Hälfte der *Alten Prager Akten* – eine vergleichbare Verzeichnung vorliegt.³³ Schließlich wäre auch die Aufnahme der überlieferten Prozessakten des als Höchstgericht für die schwedischen Reichsstände tätigen Wismarer Tribunals in Erwägung zu ziehen. Für dessen gegenüber Reichskammergericht und Reichshofrat bei weitem geringere Überlieferung (circa 5 000 Verfahren)³⁴ hat sich in jüngster Zeit durch das 2003 begonnene Verzeichnungsprojekt der Erschließungsstand wesentlich verbessert.³⁵ Da sowohl die Verzeichnung der Prozesse des *Reichshofrats* in Wien als auch der des *Wismarer Tribunals* in mehreren norddeutschen Verwahrarchiven³⁶ in enger Anlehnung an die DFG-Richtlinien erfolgt,³⁷ kann von einem grundsätzlich vergleichbaren Informationsgehalt wie bei den Inventaren des Reichskammergerichts ausgegangen werden.³⁸

Die Erfassung der Daten geht von dem Grundprinzip aus, dass die formalen Daten wie etwa Parteinaamen, Prokuratoren oder auch Vorinstanzen, unmittelbar aus den Inventaren übernommen werden. Derartige In-

³³ Nach momentanem Stand sind damit circa 3 000 Reichshofratsprozesse entsprechend den DFG-Richtlinien inventarisiert.

³⁴ Nach der Schätzung von Nils Jörn wären darüber hinaus 10 000 bis 12 000 Prozesse rekonstruierbar; vgl. Nils Jörn: *Das Archiv des Wismarer Tribunals*. In: *Integration durch Recht. Das Wismarer Tribunal (1653–1806)*. Hg. von Nils Jörn, Bernhard Diestelkamp und Kjell Åke Modéer (Quellen und Forschungen zur Höchsten Gerichtsbarkeit im Alten Reich 47). Köln, Weimar, Wien 2003. S. 329–366, S. 362.

³⁵ Gegenwärtig sind circa 1600 Prozesse inventarisiert.

³⁶ In den Stadtarchiven Wismar und Greifswald sowie dem Staatsarchiv Stade.

³⁷ Zu den Modifizierungen bei der Verzeichnung der Prozessakten des Wismarer Tribunals vgl. auch Hans-Konrad Stein: *Bericht über den Tribunalsbestand im Stadtarchiv Wismar und Vorschläge zur Verzeichnung der Tribunalsakten*. In: *Integration durch Recht*, wie Anm. 34, S. 567–570.

³⁸ Im Übrigen sei angemerkt, dass das bei der Reichskammergerichts-Datenbank benutzte Grundmodell auch für die Erfassung der Rechtsprechung territorialer Obergerichte anwendbar sein dürfte.

dividualeingaben erfolgen nur dort, wo eine Formalisierung nicht möglich oder sinnvoll ist. Sie sind per se fehleranfällig, was eine ständige, zeitaufwendige Kontrolle erforderlich macht. Allerdings schlägt auch hier der Vorteil einer Datenbank gegenüber Druckversionen durch – nämlich die Möglichkeit der beliebigen Veränderbarkeit, die selbstverständlich auch Fehlerkorrekturen einschließt. Allerdings können auch bei diesen formalen Daten Vereinheitlichungen sinnvoll sein. Das gilt in besonderem Maße für die Erfassung der Prokuratoren.³⁹

Demgegenüber werden insbesondere aus den Sachverhaltsbeschreibungen und den Darin-Vermerken entnommene Informationen mit vergleichbaren Inhalten in einem formalisierten System erfasst. Dabei gilt grundsätzlich, dass so weitgehend wie möglich formalisiert wird. Dafür gibt es zwei Varianten: Erstens und hauptsächlich erfolgt die Dateneingabe anhand vorformulierter aber jederzeit veränderbarer mehrstufiger Auswahllisten, bei denen eine hohe Fehlerresistenz besteht. Diese Auswahllisten sind nach dem Prinzip der Differenzierung in ‚Allgemeines‘ und ‚Besonderes‘ aufgebaut, wobei bei Abfragen der betreffende allgemeine Oberbegriff die Summe der besonderen Unterbegriffe jeweils einschließt.⁴⁰

Zweitens werden bei der Dateneingabe auch sich selbst ergänzende Kombinationsfelder verwendet. Da diese Felder Schreibfehler tolerieren, erfordern sie einen erheblichen Kontrollaufwand und verursachen damit eine entsprechende Nacharbeit.

Nach gegenwärtigem Stand sind circa 25 000 Verfahrensakten des Reichskammergerichts nach den geschilderten Kriterien in der Datenbank erfasst.⁴¹ Dabei sind zurzeit folgende Bestände vollständig in die Datenbank eingegeben:

³⁹ Die Schreibweise der Namen variiert zwischen den einzelnen Inventaren, was nicht nur die Abfragemöglichkeiten beträchtlich einschränkt, sondern auch erhebliche Zuordnungsprobleme aufwirft.

⁴⁰ Zur Systematik vgl. die Anlagen 1: Verfahrensgegenstand, 2: Beweismittel und 3: Lebensbereich.

⁴¹ Zum jeweils aktuellen Stand der Datenerfassung vgl. Adresse: <http://www.ruhr-uni-bochum.de/rkg-forschung/>.

- Herzogtum Braunschweig (670 Verfahren),⁴²
- Herzogtum Nassau ohne das ehemalige Amt Reichelsheim in der Wetterau (1912 Verfahren),⁴³
- Fürstentum Lippe-Detmold (829 Verfahren),⁴⁴
- Fürstentum Schaumburg-Lippe bzw. die sich in hessischem Besitz befindliche Grafschaft Schaumburg (241 Verfahren),⁴⁵
- Grafschaft Waldeck (166 Verfahren),⁴⁶
- Großherzogtum Hessen-Darmstadt (464 Verfahren),⁴⁷
- Landgrafschaft Hessen-Homburg (102 Verfahren),⁴⁸
- Herzogtümer Holstein und Lauenburg (611 Verfahren),⁴⁹
- Prozesse von außerhalb des Deutschen Bundes gelegenen Parteien und Prozesse zwischen den Souveränen (628 Verfahren).⁵⁰

⁴² Walter Deeters: Findbuch zum Bestand Reichskammergericht und Reichshofrat 1489–1806 (Veröffentlichungen der Niedersächsischen Archivverwaltung. Inventare und kleinere Schriften des Staatsarchivs in Wolfenbüttel 2, Inventar der Akten des Reichskammergerichts 3). Göttingen 1981.

⁴³ Claudia Helm und Jost Hausmann: Repertorien des Hessischen Hauptstaatsarchivs Wiesbaden. Abteilung 1. Reichskammergericht 1. Nassauische Prozessakten, 3 Bde., Bd. 1 A–M, Bd. 2 N–Z, Bd. 3 Anhang, Indices (Veröffentlichungen der Historischen Kommission für Nassau 43, Inventar der Akten des Reichskammergerichts 12). Wiesbaden 1987.

⁴⁴ Margarete Bruckhaus und Wolfgang Bender: Inventar der Lippischen Reichskammergerichtsakten. 2 Teile, Teil 1: A–L, Teil 2: M–Z, Indices (Veröffentlichungen der staatlichen Archive des Landes Nordrhein-Westfalen, Reihe A: Inventare staatlicher Archive. Das Staatsarchiv Detmold und seine Bestände 2, Inventar der Akten des Reichskammergerichts 24). Detmold 1997.

⁴⁵ Hans-Heinrich Ebeling: Findbuch zu den Reichskammergerichtsakten 1551–1806. Bestände L 24 und H 24 (Veröffentlichungen der Niedersächsischen Archivverwaltung. Inventare und kleine Schriften des Staatsarchivs in Bückeburg 1, Inventar der Akten des Reichskammergerichts 9). Rinteln 1985.

⁴⁶ Andrea Korte: Bestand 140: Waldeckische Reichskammergerichtsakten (Repertorien des Hessischen Staatsarchivs Marburg, Inventar der Akten des Reichskammergerichts 6). Marburg 1983.

⁴⁷ Andrea Korte-Böger und Cornelia Rösner-Hausmann: Reichskammergerichtsakten, wie Anm. 16. Darmstadt 1990.

⁴⁸ Jost Hausmann: Repertorien des Hessischen Hauptstaatsarchivs Wiesbaden. Abteilung 1 Reichskammergericht. Prozessakten der Landgrafschaft Hessen-Homburg (Inventar der Akten des Reichskammergerichts 7). Wiesbaden 1984.

⁴⁹ Hans-Konrad Stein-Stegemann: Findbuch der Reichskammergerichtsakten. Abt. 390 und andere. 2 Bde., Bd. 1 Titelaufnahmen, Bd. 2 Indices (Veröffentlichungen des Schleswig-Holsteinischen Landesarchivs, 16 und 17, Inventar der Akten des Reichskammergerichts 10). Schleswig 1986.

⁵⁰ Ursula Hüllbusch und Hans Schenk: Findbücher zu Beständen des Bundesarchivs 46. Reichskammergericht. Bestand AR 1. Prozessakten (Inventar der Akten des Reichskammergerichts 18). Koblenz 1994. – Otto Koser: Repertorium der Akten des Reichskammergerichts, untrennbarer

Abgeschlossen ist auch die Eingabe der Prozessakten für die in separaten Inventaren erfassten Reichsstädte:⁵¹

- Frankfurt am Main (1667 Verfahren),⁵²
- Köln (1864 Verfahren),⁵³
- Hamburg (1386 Verfahren),⁵⁴
- Lübeck (775 Verfahren),⁵⁵
- Bremen (463 Verfahren),⁵⁶
- Dortmund (279 Verfahren; verwahrt im Staatsarchiv Münster),⁵⁷
- Wetzlar (193 Verfahren).⁵⁸

Bestand, Band 1: Prozessakten aus der Schweiz, Italien, den Niederlanden und dem Baltikum, sowie der freiwilligen Gerichtsbarkeit (Veröffentlichungen des Gesamtvereins der deutschen Geschichts- und Altertumsvereine). Heppenheim 1933. – Ders.: Repertorium der Akten des Reichskammergerichts, untrennbarer Bestand, Band 2: Prozessakten aus dem Elsass, aus Lothringen und den angrenzenden ehemaligen Reichslanden (Veröffentlichungen des Gesamtvereins der deutschen Geschichts- und Altertumsvereine). Heppenheim 1936.

⁵¹ Lediglich für den circa 1600 Prozesse enthaltenden Bestand der alten Reichsstadt Aachen liegt ein entsprechendes Inventar zurzeit noch nicht vor.

⁵² Inge Kaltwasser: Inventar der Akten des Reichskammergerichts 1495–1806. Frankfurter Bestand (Veröffentlichungen der Frankfurter Historischen Kommission XXI, Inventar der Akten des Reichskammergerichts 27). Frankfurt/Main 2000.

⁵³ Matthias Kordes: Reichskammergericht Köln (Mitteilungen aus dem Stadtarchiv von Köln, Bd. 1 Nr. 1–600 (A–F), Bd. 2 Nr. 601–1232 (G–M), Bd. 3 Nr. 1233–1677 (N–S), Bd. 4 Nr. 1678–1863 (T–Z), Inventar der Akten des Reichskammergerichts 26). Köln, Weimar, Wien 1998–2002.

⁵⁴ Hans-Konrad Stein-Stegemann: Findbuch der Reichskammergerichtsakten im Staatsarchiv Hamburg. Teil 1 Titelaufnahmen A–H, Teil 2 Titelaufnahmen J–R, Teil 3 Titelaufnahmen S–Z, Nachträge, Teil 4 Indices (Veröffentlichungen aus dem Staatsarchiv der Freien und Hansestadt Hamburg, Bd. XIII, Inventar der Akten des Reichskammergerichts 21). Hamburg 1993–1995.

⁵⁵ Hans-Konrad Stein-Stegemann: Findbuch der Reichskammergerichtsakten im Archiv der Hansestadt Lübeck. 2 Bde., Bd. 1 Titelaufnahmen A–R, Bd. 2 Titelaufnahmen S–Z, Indices (Veröffentlichungen des Schleswig-holsteinischen Landesarchivs, Bde. 18 und 19, Inventar der Akten des Reichskammergerichts 13). Schleswig 1987.

⁵⁶ Andreas Röpcke und Angelika Bischoff: Inventar der Bremer Reichskammergerichtsakten (Kleine Schriften des Staatsarchivs Bremen 22, Inventar der Akten des Reichskammergerichts 22). Bremen 1995.

⁵⁷ Günter Aders: Dortmunder Prozesse vor dem ehemaligen Reichskammergericht (Beiträge zur Geschichte Dortmunds und der Grafschaft Mark LIX. Hg. vom Historischen Verein für Dortmund und die Grafschaft Mark). Dortmund 1962.

⁵⁸ Jost Hausmann: Repertorien des Hessischen Hauptstaatsarchivs Wiesbaden, Abteilung 1 Reichskammergericht, Teil 3 Prozessakten des preußischen Kreises und der Stadt Wetzlar (3 Bde.), 3. Bd. Stadt Wetzlar, Indices (Inventar der Akten des Reichskammergerichts 8). Wiesbaden 1984–1986.

Die Erfassungsarbeiten für diese Bestände erfolgte im Rahmen eines speziellen ‚Reichsstadtprojekts‘⁵⁹ zum überwiegenden Teil drittmittelfinanziert durch die Alfred Krupp von Bohlen und Halbach-Stiftung und den Verein zur Förderung der Rechtswissenschaft e.V. Bochum.⁶⁰ Die zahlreichen, in verschiedenen territorialstaatlich geordneten Inventaren überlieferten Prozessakten reichsstädtischer Provenienz (etwa Augsburg und Nürnberg im Bayerischen Hauptstaatsarchiv München oder Ulm und Schwäbisch Hall im Hauptstaatsarchiv Stuttgart) werden im Zusammenhang mit der Erfassung der entsprechenden Inventare als Prozesse reichsstädtischer Herkunft gesondert gekennzeichnet und auf diese Weise auch einzeln als Reichsstädte abfragefähig gemacht. Damit sind nunmehr circa 7 100 Verfahren reichsstädtischer Provenienz in der Datenbank verfügbar.

In jüngster Zeit wird mir zunehmend die durchaus berechtigte Frage gestellt, wann, wo und unter welchen Voraussetzungen die *RKG-Datenbank* allgemein benutzbar sein wird. Dazu läßt sich heute und hier zunächst Folgendes sagen: Zurzeit steht nur eine Access-Version der Datenbank auf CD-ROM zur Verfügung. Insoweit konnten bisherige Interessenten lediglich durch konkrete Nachfragen zu Einzelproblemen die Datenbank nutzen.

Gegenwärtig arbeiten wir intensiv daran, im Spätsommer bzw. im Herbst diesen Jahres die *RKG-Datenbank* ins Internet zu stellen. Das wird notwendigerweise verbunden sein mit der gleichzeitigen Veröffentlichung mindestens einer Kurzanleitung für potentielle Nutzer, die in das komplexe System der Datenbank und deren Möglichkeiten einführen soll. Allerdings werden in diesem Jahr nur die vollständig erfassten Bestände veröffentlicht werden. Das hängt damit zusammen, dass für unvollständig in die Datenbank eingegebene Bestände die notwendigen Feinarbeiten und Korrekturen noch nicht geleistet werden können.

Allerdings werden bereits im nächsten Jahr die beiden großen Bestände in Düsseldorf (für den Regierungsbezirk Düsseldorf der preußischen Rheinprovinz) und Münster (für die preußische Provinz Westfalen) mit zusammen fast 13 000 Verfahren vollständig in der Datenbank erfasst und

⁵⁹ Erste Ergebnisse dieses Teilprojekts haben zwei Mitarbeiter – Dr. Kathrin Dirr-Jansen und Torsten Joecker – auf einer Arbeitstagung der Gesellschaft für Reichskammergerichtsforschung vorgestellt; zur Schriftfassung dieses Beitrages vgl. Anm. 29. Ferner ist in diesem Zusammenhang auch die Dissertation von Kathrin Dirr-Jansen zu nennen (wie Anm. 26).

⁶⁰ Beiden Institutionen sei auch an dieser Stelle recht herzlich gedankt.

entsprechend bearbeitet sein. Damit würde sich Ende 2007 der für die allgemeine Nutzung zugängliche Datenbestand nahezu verdoppeln.

Anlage 1: Verfahrensgegenstand⁶¹

Oberbegriffe

1. Ehe/Familie
2. Einwendungen/Einreden
3. Erbrecht
4. Feudalrecht
5. Iniurien
6. Konkursrecht
7. Landfriedensbruch
8. Personenrecht
9. Policity
10. Recht/Gericht
11. Sachenrecht
12. Schuldrecht
13. Steuern/Abgaben
14. Stiftungsrecht
15. Strafrechtliche Bezüge
16. Verfahren
17. Verfassung/Verwaltung
18. Vergleich

Gliederungsschema (Beispiel)

1. Ehe/Familie
2. Ehe/Familie † Ehe
3. Ehe/Familie † Ehe † Ehegüterrecht
4. Ehe/Familie † Ehe † Ehegüterrecht † Ehevertrag

⁶¹ Zurzeit stehen in 18 Obergruppen insgesamt 488 einzelne Auswahlmöglichkeiten zur Verfügung.

Anlage 2: Beweismittel

1. Augenschein
2. KA
3. Parteienvernehmung
4. Sachverständige
5. Sachverständige † Juristenfakultät
6. Sachverständige † Schöffenstuhl
7. Urkunden
8. Urkunden † Abschied
9. Urkunden † Freiheitsbrief
10. Urkunden † Königsurkunde
11. Urkunden † Lehnsurkunde
12. Urkunden † Privileg
13. Urkunden † Rechtsquellen
14. Urkunden † Rechtsquellen † Dorfordnung
15. Urkunden † Rechtsquellen † Gerichtsordnung
16. Urkunden † Rechtsquellen † Landesordnung
17. Urkunden † Rechtsquellen † Weistum
18. Urkunden † Testament/Vermächtnis/Legat/Kodizill
19. Urkunden † Urteil
20. Urkunden † Vertrag/Rezess
21. Urkunden † Zinsregister
22. Verweis auf Acta priora
23. Zeugen

Anlage 3: Lebensbereich

1. Archivwesen
2. Armenwesen
3. Bauangelegenheiten
4. Bauerngüter/Rittergüter
5. Bergbau
6. Deutschritterorden
7. Dominikanerorden
8. Fabriken/Manufakturen
9. Familienangelegenheiten

10. Forstwesen
11. Geldwesen
12. Gemeindeangelegenheiten
13. Genealogie
14. Gesundheitswesen
15. Handel und Gewerbe
16. Handwerk
17. Jahrmärkte
18. Johanniterorden
19. Juden
20. KA
21. Kaiser/Reichsangelegenheiten
22. Kirchenangelegenheiten
23. Klöster und Stifte
24. Kreisangelegenheiten
25. Kriminalität
26. Kurfürsten
27. Liegenschaftswesen
28. Markangelegenheiten
29. Militärwesen
30. Mühlensachen
31. Nur Vorakte
32. Obrigkeit
33. Rechnungswesen
34. Religionsangelegenheiten/Reformation
35. Schifffahrtswesen
36. Schulwesen
37. Staat/Verwaltung/Justiz
38. Stiftungswesen
39. unklar
40. Vermögensangelegenheiten
41. Wappen/Siegel
42. Zunftwesen

Anlage 4

RGK-Verfahren - [F_Main : Formular]

Suche | Verfahren | Zuständigkeit/Verfahrensart/ Streitwert/ Instanz/ Verfahrensausgang/ Beweismittel | Inhalt | Beschreibung/ Bemerkungen

Sachfragen

Verfahrensgegenstand Ehe/Familie | Ehe | Ehegüterrecht | Witwenrecht
 Erbrecht | Erbengemeinschaft | Erbteilung

Ver-fegenst-/ Erläuterung

Lebensbereich Familienangelegenheiten

Verfahrensausgang

Beweismittel Urkunden | Lehnurkunde
 Urkunden | Urteil

Beweismittel/ Erläuterung

Zuständigkeit

Verfahrensart

Streitwert

Armenrecht

Formale Daten

Instanz 1 RKG

Hoheitsträger

Reichskreise (Instanzen)

Verfügbarkeit der Verfahrensakten

Verweis

Personendaten

GR Buchstabe GR Nummer

Namen der Parteien

Status im Verfahren

Status in der Vorinstanz

Beruf

Stand

Sozialstatus Adel

Art der Vertretung

Name des Vertreters

Titel des Vertreters

Mandatsverhältnis

Mehrfach

Jahr der Bevollmächtigung

ortsbezogene Daten

Hamburg

Städte

Sitz der Parteien

Reichskreis Niedersachsen

Zeitliche Daten

Beginn des Verfahrens 1601 1700 Verfahrensdauer

Ende des Verfahrens 1601 1900

Datensatz: 1 von 35846

Art des Gerichts lt. Auswahlliste, bei RKG = Entfällt

Start | F:RGK_WECHSEL | RKG-Verfahren ... | 14:04

Konzepte zur Bereitstellung digitalisierter frühneuzeitlicher Quellen¹

Thomas Stäcker

Die Medienkonversion des gedruckten kulturellen Erbes des deutschen Sprachraums in eine digitale Form gehört zu den großen Herausforderungen unserer Zeit. Kaum jemand bestreitet heute noch, dass die digitale Bereitstellung unserer Kulturgüter ein wichtiger Faktor für deren Rezeption ist. Was nicht im Web ist, ist nicht in der Welt. Doch geht es dabei nicht nur um die bloße Bereitstellung. Mit der verstärkten Verlagerung auch geistes- und kulturwissenschaftlicher Kommunikation und Publikation in das Internet bildet sich allmählich eine neue Forschungspraxis heraus. Das Internet entwickelt sich zu einem integralen Forschungsraum, der Bibliographien, Datenbanken, Suchmaschinen und Kataloge ebenso umfasst wie Forschungs- bzw. Sekundärliteratur und digitalisierte Quellenwerke. Ressourcen, die sich nicht in diesen Forschungsraum integrieren lassen, werden es in Zukunft schwer haben, noch wahrgenommen und wissenschaftlich rezipiert zu werden. Umgekehrt verändert sich auch das kultur- und geisteswissenschaftliche Arbeiten und passt sich immer stärker den neuen Medienbedingungen an. Ein grundlegendender Wandel bahnt sich an, wie er heute schon vielerorts bei den Technik- und Naturwissenschaften zu beobachten ist: Ohne Medienbruch gelangt man von der Quelle direkt zur Literatur, die Literatur vernetzt sich untereinander, von der Literatur kommt man wiederum zu Quellenwerken. Elektronisch lesbare Kataloge und Bibliographien bzw. Datenbanken dienen als wissenschaftlich verlässliche Drehscheiben dieser Entwicklung.

¹ Für die Druckfassung des Vortrages wurden einige Angaben ergänzt und revidiert, der Vortragscharakter aber im Großen und Ganzen beibehalten.

Unter dem Eindruck von ‚Google Print‘², das viele europäische Politiker und Kulturverantwortliche aufschreckte, hat sich das Tempo dieser Entwicklung beschleunigt und auf nationaler wie europäischer internationaler Ebene diverse Initiativen ausgelöst.

Die jetzt aufgekommene Digitalisierungseuphorie lässt indes leicht vergessen, dass das Herstellen von Digitalisaten nur die eine, mittlerweile vermutlich sogar die unproblematischste Seite der Medaille ist. Denn das schönste Digitalisat nutzt nichts, wenn es nicht gefunden wird. Ein zentrales Nachweisinstrument fehlte jedoch bzw. blieb ein Desiderat.³ In der von Manfred Thaller betreuten ‚RetroEval-Studie‘⁴ ergab sich bei den Benutzern ein erhebliches Informationsdefizit. Zahlreiche Befragte waren überrascht, wie viel schon digitalisiert ist, die wenigsten wussten, wo die für sie relevanten Altbestandsquellen zu finden sind. Wie konnte es dahin kommen? Die Deutsche Forschungsgemeinschaft hat im Programm *Retrospektive Digitalisierung von Bibliotheksbeständen* seit 1997 eine Fülle von Projekten gefördert, die zum Ziel hatten, den Direktzugriff auf für Forschung und Lehre wichtige Bestände, den Mehrfachzugriff auf viel genutzte Literatur, die digitale Bereitstellung schwer zugänglicher Bestände und die erweiterte Nutzung bisher nur wenig bekannter Materialien verbessern zu helfen. Diese Ziele konnten zwar in bemerkenswert vielen Einzelprojekten erreicht werden, doch lag die Kommunikation der Projekte untereinander und die Bündelung der Anstrengungen in einer gemeinsamen virtuellen Bibliothek trotz des programmatischen Namens *Verteilte Digitale Forschungsbibliothek* lange Zeit nicht im Brennpunkt der Aufmerksamkeit. Auch die für die Koordination und Beratung der Digitalisierungsaktivitäten seinerzeit in Göttingen und München aus der Taufe gehobenen Digitalisierungszentren erfüllten diesbezüglich nicht die in sie gesetzten Erwartungen und konzentrierten sich – durchaus erfolgreich – auf eigene Projekte.⁵ Als Konsequenz dieser Entwicklung entstand eine buntscheckige Landschaft von digitalen

² Adresse: <http://books.google.com/>.

³ Vgl. auch Thomas Stäcker: Das ist doch alles im Netz! – Angebot und Nutzen von digitalisierten Altbestandsquellen im Internet. Vortrag gehalten auf dem Bibliothekartag in Düsseldorf 2005. Adresse: <http://www.opus-bayern.de/bib-info/volltexte/2005/73/>.

⁴ Manfred Thaller und andere: „Retrospektive Digitalisierung von Bibliotheksbeständen“ – Evaluierungsbericht über einen Förderschwerpunkt der DFG. Universität Köln 2005. http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/retro_digitalisierung_eval_050406.pdf (letzte Einsichtnahme am 18.11.2005).

⁵ Ebd., S. 11.

Sammlungen auf unterschiedlichen technischen Plattformen ohne einheitliche Standards. So ist es nur zu verständlich, dass heutige Nutzer, wie in der *RetroEval*-Studie festgestellt, nach einem zentralen Portal rufen und in den Äußerungen der Studie allenthalben ein Bedürfnis nach allgemeinen Standards in der Erschließung und Präsentation erkennbar ist.

Um diesem gravierenden Mangel abzuhelpfen, fördert die DFG seit 2004 das *Zentrale Verzeichnis Digitalisierter Drucke* (zvdd). Neben dem Nachweis von Drucken gibt es aber auch noch andere Portale. Hier eine Übersicht über die wichtigsten nach Materialarten:

- Mittelalterliche und frühneuzeitliche Handschriften: *manuscripta mediaevalia*,
- Autographen und Nachlassmaterialien: *Kalliope*,
- Bundesarchiv,
- Inkunabeln: INKA, vdIb, ISTC,
- Drucke bzw. Zeitschriften: *Zentrales Verzeichnis Digitalisierter Drucke* (zvdd),
- Bildarchive bzw. Kunstobjekte: *Bildarchiv Foto Marburg*.

Mit den zentralen Portalen wie *Manuscripta mediaevalia*, *Kalliope*, *zvdd* oder auch, von fachwissenschaftlicher Seite aus durch die *Virtuellen Fachbibliotheken* (Vifa) wird die Basis gelegt für übergreifende Suchen. Der konzeptionell richtige Weg ist die präzise materialbezogene Beschreibung und erst in einem weiteren Schritt die materialunabhängige Abstraktion. Qualität und Quantität der Information stehen in einem umgekehrt proportionalen Verhältnis zueinander. In einem lokalen Projekt wird man die spezifischsten Informationen finden, in Google die allgemeinsten. Dazwischen liegen die material- und fachspezifischen Portale, die wissenschaftliche von nicht-wissenschaftlicher Information und Quellennachweisen trennen, die formal oder inhaltlich von der Materialbasis abstrahieren. Ich schicke diesen zentralen Aspekt vorweg, um deutlich zu machen, dass jede Form der Digitalisierung ohne einen zentralen Nachweis verfehlt ist. Zumindest ein Eintrag im Katalog der Bibliothek muss verpflichtend sein.

Ist das Digitalisat katalogisiert und der zentrale Nachweis gegeben: Welche Potentiale bietet die Digitalisierung frühneuzeitlicher Quellen. Wenn ich mich hier auf frühneuzeitliche Ressourcen beschränke, so will ich keiner speziellen Form der Digitalisierung für frühneuzeitliche Quellen das

Wort reden. Ich beschränke mich vielmehr auf Beispiele aus dem Bereich der frühen Neuzeit. Mutatis mutandis sind meine Ausführungen natürlich auch auf andere Materialien übertragbar. Welches sind also die aus meiner Sicht zentralen Entwicklungslinien für den bibliothekarischen und wissenschaftlichen Umgang mit digitalen Ressourcen aus der frühen Neuzeit? Ich sehe hier vor allem drei Themenfelder:

- Vernetzung
- Kooperative Arbeit und ‚Content Syndication‘
- Elektronische Edition

1. Vernetzung

Gelegentlich fragt man sich, ob es eigentlich sinnvoll ist, eine Quelle zu digitalisieren, die bereits als Reprint vorliegt. Natürlich hat man als Privatperson oft nicht die Mittel, sich teure Reprints zu kaufen und natürlich ist es vorteilhaft, wenn man von überall her kostenfrei auf digitale Ressourcen zugreifen kann. Doch genügt das als Grund für die Digitalisierung? An diese Überlegung knüpft sich die Frage, welche Rolle das Internet in Zukunft für die kultur- und geisteswissenschaftliche Forschung spielen wird. Bislang ist die Digitalisierung noch zu sehr vom Gedanken der Sekundärform geprägt. Das Digitalisat wird als eine neben anderen Reproduktionen angesehen, wie ein Microfiche, -film, Papierkopie oder sonstige analoge Kopie. Darin wird jedoch das Potential eines Digitalisats unterschätzt, das vor allem darin beruht, im Web angeboten werden zu können. Eine digitale Kopie auf CD-ROM ist in der Tat im Prinzip nichts anderes als eine Papierkopie, im Netz angeboten, entfalten sich dank der Hypertextualität des Mediums jedoch andere Wirksamkeiten. Vor allem darin, in der Möglichkeit, Inhalte vernetzen zu können, so meine Behauptung, liegt der wissenschaftliche Mehrwert der Digitalisierung, weit mehr als in einer bequemer zugänglichen Sekundärform.

Wenn die Forschung, auch die geistes- und kulturwissenschaftliche, in Zukunft selber stärker das Internet als Publikationsmedium nutzen sollte, und dafür sprechen eine Reihe von Indizien, dann liegen die Vorteile der Vernetzung auf der Hand. Die Fußnote wird zum Link. Benutzte Quellen wie auch Sekundärliteratur lassen sich umgehend überprüfen. Kataloge

und wissenschaftliche Suchmaschinen fungieren als zuverlässige Drehscheiben dieser Vernetzung. Gerade die Möglichkeit, Quellen medienintegral in eine wissenschaftliche Arbeit einbeziehen zu können, ist eine attraktive Perspektive und ein wichtiger Antrieb für wohl alle derzeit unternommenen Digitalisierungsprojekte.

Dies ist oft mit geringem Aufwand zu erreichen, wie sich am Wolfenbüttler Projekt *Leibnizressourcen Digital*⁶ zeigt. In diesem Kooperationsprojekt mit der Berlin Brandenburgischen Akademie der Wissenschaften wurden rund 120 Titel in Wolfenbüttel digitalisiert und mit der Online-Edition der *Naturwissenschaftlich-medizinisch-technischen Schriften* von Leibniz verknüpft. Wo Leibniz einen Titel bzw. eine Passage zitiert, zum Beispiel aus den *Experimenta nova* des Otto von Guericke (Amsterdam 1672)⁷, kann man die fragliche Stelle seitengenau anspringen und das Buch gleichsam an dieser Stelle aufblättern. Umgekehrt führt ein Button „citations“ vom Angebot der Wolfenbütteler Seite über eine Google-Suche nach Ressourcen, die den Link zitieren, unter anderem zu der Seite der BBAW, so dass beide Angebote bidirektional verlinkt sind.

Voraussetzung für die Vernetzung bzw. Zitierfähigkeit ist ein stabiler Link: Die schon genannte PURL (Persistent URL), oder aber ein URN, der über einen Resolver funktioniert. In Wolfenbüttel gibt es hierfür verschiedene Lösungen, die im Ausgang von der Signatur diverse Systeme bedienen (vergleiche Abbildung 1).

Da die URN in Wolfenbüttel aus Aufwandsgründen – es müsste sonst bei der Deutschen Bibliothek für jede Seite eine eigene URN angelegt werden – nur auf Titelebene angewendet wird, wir aber für das Projekt die Seitenebene benötigen, arbeiten wir für die seitengenaue Verlinkung nur mit PURLs. Entscheidend ist aber weniger, wie es gemacht wird, wichtiger ist vielmehr, dass man überhaupt eine persistente Verlinkung mit eindeutigem Identifier für die Objekte vorsieht. An dem Beispiel wird auch deutlich, warum Wolfenbüttel neben Gründen der Authentizität relativ reserviert Sekundärarchiven im Netz gegenübersteht. Denn Ressourcen, die diese Quellen im Netz zitieren, werden von Wolfenbütteler Seite aus nicht mehr ohne weiteres gefunden.

⁶ Adresse: <http://www.hab.de/forschung/projekte/leibnizressourcen.htm>.

⁷ Adresse: <http://diglib.hab.de/drucke/34-5-phys-2f/start.htm?image=00005>.

Signatur des Originals: Einbl. Xb FM 28

Signatur des Digitalisats: `drucke/einbl-xb-fm-28` (Grundbestandteil des Identifiers)

URL: <http://diglib.hab.de/wdb.php?dir=drucke/einbl-xb-fm-28>
(seit 2004; früher wie PURL)

PURL: <http://diglib.hab.de/drucke/einbl-xb-fm-28/start.htm>
(seit 1998)

URN: `urn:nbn:de:gbv:23-drucke/kb-53-2f-25`
(seit 2005)

Hauseigener Resolver für die URN (und andere Identifier):
<http://diglib.hab.de/?um=urn:nbn:de:gbv:23-drucke/einbl-xb-fm-285>

Resolver der Deutschen Bibliothek:
<http://nbn-resolving.de/urnresolver.pl?um=urn:nbn:de:gbv:23-drucke/einbl-xb-fm-285>

Abbildung 1

2. Kooperative Arbeit und Content Syndication

Das Internet erlaubt neue Kooperationsszenarien in der Bereitstellung digitaler Quellen. Kooperationen lassen sich in quantitative, qualitative und technische unterscheiden. Während es bei der quantitativen Kooperation um die gemeinsame Bereitstellung von gleichartigen Objekten geht, zielt die qualitative auf die differenzierte und arbeitsteilige Bearbeitung verschiedener Schritte am selben Objekt. Technische Kooperationen wiederum verteilen inhaltliche und datenbanktechnische Prozesse auf verschiedene Partner oder gliedern einzelne technische Aspekte aus einem Angebot aus bzw. ‚sourcen‘ sie aus.

Als Beispiel zu einer quantitativen Kooperation möchte ich das Projekt *vdIb*⁸ nennen. Die Universität Köln, die Stadt- und Universitätsbibliothek Köln und die Herzog August Bibliothek haben zusammen mehr als 1 300 Inkunabeln und knapp 300 000 Seiten digitalisiert und inhaltlich erschlossen bzw. mit Strukturdaten zur Navigation versehen, wobei sie pragmatisch bei 1485 eine Zeitgrenze gezogen haben. Titel, die älter als dieses Datum waren, wurden von Köln, jüngere von Wolfenbüttel digitalisiert. Obwohl Köln wie Wolfenbüttel die Quellen auf dem eigenen Server publizieren, werden doch beide unter einer Oberfläche zusammengefasst. Zu-

⁸ Adresse: <http://inkunabeln.ub.uni-koeln.de/>.

gleich werden zum Beispiel die Ressourcen aus Wolfenbüttel über den OPAC der Bibliothek, daher mittelbar auch über den Verbund und den KVK, das Inkunabelportal *INKA*, den *GW* und den, allerdings noch nicht freigeschalteten, *ISTC* gefunden.

So werden zum einen Inhalte bei distributiver Datenhaltung kooperativ zu für die frühneuzeitliche Forschung nützlichen Corpora aggregiert, zum anderen erlaubt die elektronische Datenhaltung Content Syndication über verschiedene Oberflächen und Angebote. Noch funktioniert es gewissermaßen zu Fuß. Daten werden in meist proprietären Formaten übertragen und eingearbeitet. Mit der weiteren Verbreitung von OAI⁹ – eine entsprechende Schnittstelle wird derzeit an der HAB implementiert – und verbindlichen Austauschformaten, werden sich diese Prozesse automatisieren. Äußerst effektiv funktioniert das heute schon in der Blogger-Szene, die vor allem über RSS-Feeds Daten aggregiert und erfolgreich Content Syndication betreibt.

Qualitative Kooperationen sind nach meinem Eindruck noch selten. Gerade sie sind aber als Szenarien einer engeren Zusammenarbeit von Wissenschaft und Bibliothek, von frühneuzeitlicher Forschung und Quellenbereitstellung, besonders interessant. Ein kleines Beispiel sei die externe Erschließung eines Wolfenbütteler Emblembuchs in einem niederländischen Projekt zur Liebesemblemik.¹⁰ Dort wurden zahlreiche Einzelembleme mit ICON-CLASS-Notationen¹¹ versehen und so differenziert erschlossen. Von den angebotenen Thumbnails gelangt man direkt zum Wolfenbüttler Digitalisat.

Eine qualitative Zusammenarbeit findet auch mit dem Projekt *Camena* bzw. dem *Thesaurus eruditionis* statt¹². Vereinbart ist, Drucke, die bei *Camena* nicht vorhanden oder nicht digitalisierbar sind, in Wolfenbüttel zu digitalisieren, des weiteren sollen einige dieser Drucke in Heidelberg im Volltext erfasst und getaggt werden, von den Heidelberger Volltexten kann man dann zu den Wolfenbütteler Images springen.

Zuletzt möchte ich noch eine technische Kooperation im Projekt *Festkultur Online*¹³ vorstellen. Das Projekt ist übrigens wie *vdlb* zugleich eine quantitative Kooperation, insofern in Warwick bzw. der British Library und Wolfenbüttel Festbücher nach gleichen Richtlinien auf Seitenebene er-

⁹ Adresse: <http://www.openarchives.org/>.

¹⁰ Unter Adresse: <http://www.mnemosyne.org/mia/> mit Stichwort HAB suchen.

¹¹ Adresse: <http://www.iconclass.nl>.

¹² Adresse: <http://www.uni-mannheim.de/mateo/camenahtdocs/camenaref.html>.

¹³ Adresse: <http://www.hab.de/bibliothek/wdb/festkultur/index.htm>.

schlossen und unter einer gemeinsamen Oberfläche in Wolfenbüttel angeboten wurden. Der technische Aspekt der Kooperation gelang jedoch durch eine Zusammenarbeit mit ICONCLASS. ICONCLASS hat Wolfenbüttel einen eigenen Browser¹⁴ eingerichtet, der es erlaubt, bequem Notationen zu suchen. Ein echter Mehrwert für das Projekt.

3. Elektronische Edition

Großes Entwicklungspotential hat meines Erachtens innerhalb der Frühneuzeitforschung die elektronische Edition. Der Begriff ist hier tempertativ gebraucht, denn es ist definitorisch nicht einfach, eine elektronische oder digitale Edition in ihren Konstituentien eindeutig zu bestimmen. Man könnte natürlich argumentieren, sie enthält mindestens das, was eine gedruckte Edition bietet. Jedoch erlaubt das Internet sowohl Skalierungen und Modularisierungen als auch, wiederum bedingt durch die Hypertextualität, andere Formen der Präsentation.

Wir haben zum Beispiel in einer elektronischen Edition von Briefen Kirchers an Herzog August den Jüngeren¹⁵ versucht, das Spektrum durchzuspielen. Geboten werden auf der Basis von XML Transkription, Transkription mit einem zugegebenermaßen schmalen Apparat, ein Kommentar, der von Fletcher stammt, sowie eine Übersetzung – nicht zu vergessen das digitale Image.

In einem anderen Editionsprojekt, das in Zusammenarbeit von Jolanta Gelumbeckaite, Christian Heitzmann und mir durchgeführt wurde, hat Gelumbeckaite eine Transkription mit Fußnoten und eine Einleitung zu einer in Wolfenbüttel aufgefundenen Handschrift verfasst, Heitzmann und ich haben eine Übersetzung davon angefertigt. Die Transkription ist bereits zusammen mit den Images online¹⁶, die komplexe Kompletteedition mit Images, Übersetzung, Transkription und Einleitung wird als eigenständiges Werk in Kürze auch noch einmal separat online erscheinen. Obwohl es unsinnig erscheinen mag, werden die Images und die Transkription dort in der Tat noch einmal

¹⁴ Zum Beispiel Adresse: http://icontest.iconclass.nl/libertas/ic?task=getnotation-&datum=9&style=notationbb_hab.xml&taal=de beim Verfolgen des Links „Festkultur Online – HAB Wolfenbüttel“ gelangt man direkt in die Festkulturdatenbank und findet Abbildungen zur Mythologie.

¹⁵ Adresse: <http://diglib.hab.de/edoc/ed000005/start.htm>.

¹⁶ Adresse: <http://diglib.hab.de/mss/11-14-aug-2f/start.htm?image=00261r>.

angeboten. Wir kommen damit einerseits dem Wunsch der Deutschen Bibliothek nach der Transferfähigkeit von digitalen Werken nach, mit anderen Worten, alle Bestandteile der Edition müssen leicht auch zur DDB übertragen werden können. Zum anderen betrachten wir die Images, wenn sie in eine elektronische Edition integriert sind, als eigenständig, denn sie werden in einen anderen editorischen Kontext eingebunden und fungieren nicht mehr nur als Faksimileedition mit gegebenenfalls Transkription. Derzeit ist eine weitere Transkription und Übersetzung von Heitzmann und mir in Arbeit, nämlich Justus Lipsius *de bibliothecis*, das noch nie ins Deutsche übertragen wurde. Als digitale Faksimiles online sind die Erstausgabe von 1602¹⁷ und die für die Textüberlieferung relevante spätere Auflage von 1620¹⁸. Weniger für die Übersetzung, aber für die Transkription sollen Textabweichungen notiert werden. Auch hier wird ein stufenweiser Editionsprozess verfolgt. Zunächst das digitale Faksimile, dann gegebenenfalls die Transkription oder Strukturdaten oder aber die Übersetzung.

Was aber macht die Edition zur Edition? Im Kontext der Wolfenbütteler Bibliothek: Wann ist der Punkt gekommen, dass man von einer Faksimileedition mit gegebenenfalls Strukturdaten und Transkription zu einer eigenständigen Publikation wechselt? Ich muss zugeben, dass uns das Problem bis heute Kopfzerbrechen bereitet. Um dies auch noch an einer anderen Baustelle zu veranschaulichen:

Daniel Georg Morhofs *Polyhistor* wird nach den Images hier zitiert <<http://diglib.hab.de/drucke/ea-490/start.htm>> und als eigenständiger Text ist er zugleich vorläufig hier: <<http://diglib.hab.de/drucke/ea-490/transcript-roh.htm>> verfügbar.

Wie gesagt, vorläufig, denn es handelt sich um einen nicht korrigierten und nur notdürftig getaggten Text, der gleichwohl nützlich ist. Ich denke, dass der Editionsstatus nicht verbindlich objektivierbar ist. Mindestens vorliegen müssen nach meinem Gefühl bei Werken der frühen Neuzeit die Images und die – mit Mindestmarkup versehene – Transkription. Allerdings sollte die Publikation auch eine entweder implizite oder explizite Erklärung des Autors enthalten, dass es sich nicht mehr um ‚work in progress‘ handelt.

¹⁷ Adresse: <http://diglib.hab.de/drucke/qun-59-9-1/start.htm>.

¹⁸ Adresse: <http://diglib.hab.de/drucke/ba-474/start.htm>.

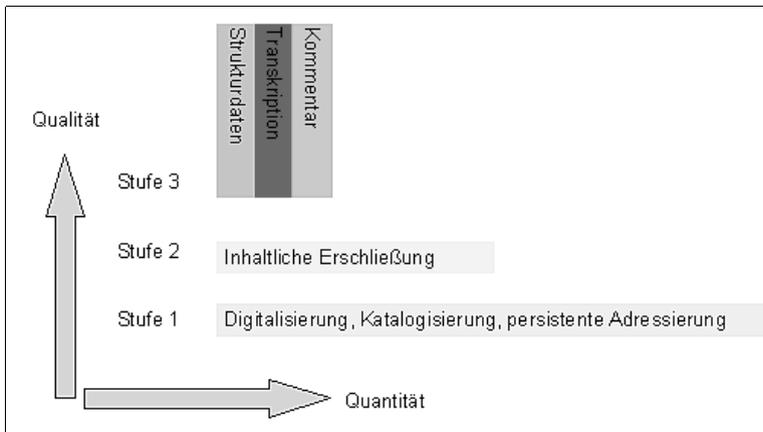


Abbildung 2

Für die Bereitstellung von frühneuzeitlichen Texten kann man die verschiedenen Stufen in einem Schema gemäß Abb. 2 visualisieren, wobei meines Erachtens nur die Aspekte der Stufe 3 eine Edition konstituieren können. Man sieht hieran auch, dass nur wenige Texte editorisch aufbereitet werden können, viele jedoch digitalisiert, etwas weniger auf Titelebene inhaltlich differenziert erschlossen. Was die Quellen der frühen Neuzeit anlangt, so sehe ich die Stufe 1 und 2 technisch gelöst und hoffe, dass man in den nächsten 5 Jahren durch massive Digitalisierungskampagnen das Angebot substantiell verbessert. Bei der Stufe 3 sehe ich noch viele konzeptionelle Probleme, für die sich aber sicher in naher Zukunft Lösungen finden werden – nicht zuletzt durch das wachsende Interesse der Universitäten, Akademien und sonstigen Forschungseinrichtungen an digitalen Editionen, ein Interesse, das nach meiner Überzeugung in naher Zukunft zu einem ‚digital shift‘ in der Editionslandschaft führen wird. Abbildung 2. Es steht zu hoffen, dass die digitale Quellenbereitstellung und -erschließung ein kooperativer Prozess sein wird, der mit Blick auf die Quantität, die Qualität und die Technik arbeitsteilig die Potentiale des Internet nutzt und die Frühneuzeitforschung, und nicht nur diese, nachhaltig zu befördern vermag.

Archive in der digitalen Welt

Informationstransfer zwischen Verwaltung und Wissenschaft

Rainer Hering

1. Einleitung

„Auf meinem Rechner habe ich 10 000 E-Mails, die können Sie gern haben“, entgegnete mir ein leitender Beamter einer Hamburger Behörde, als ich mich ihm als Archivar vorstellte und meine Aufgaben erläuterte. Dieses bereitwillige Angebot, Unterlagen ins zuständige Archiv abzugeben, ist trotz der eindeutigen Rechtslage durch die Archivgesetze keineswegs selbstverständlich und daher nur zu begrüßen. Doch hinter diesem freundlichen Satz, der genauso gut aus dem privaten oder privatwirtschaftlichen Bereich stammen könnte, verbergen sich gravierende Probleme für Archive und für die historische Forschung. Und dabei handelt es sich trotz der relativ großen Zahl von E-Mails in erster Linie nicht um quantitative Schwierigkeiten.

Bleiben wir zunächst noch bei dem Beispiel. Was können wir dem Satz „auf meinem Rechner habe ich 10 000 E-Mails, die können Sie gern haben“ entnehmen? Offenbar pflegt dieser hohe Beamte eine umfangreiche elektronische Korrespondenz, die er lokal, das heißt auf seinem Rechner oder in einem Netzwerk, nur für ihn zugänglich, abgespeichert hat. Diese Schriftwechsel sind wohl nicht ausgedruckt, mit einer Verfügung versehen und in die zuständige Sach- oder Handakte gegeben worden. Auch im weiteren Verlauf des Gespräches fielen die Begriffe Akte und Registratur nicht. Die Schreiben sind also nicht im Kontext des jeweiligen Vorgangs in einer Akte in der Zentralregistratur seines Amtes oder seiner Dienststelle zu finden, auch scheint es keine klassische Sachbearbeiterablage im Büro zu geben.

Offen bleiben muss die Anordnung dieser 10 000 E-Mails. Es kann aber vermutet werden, dass der Beamte sich selbst eine gänzlich individuelle, für seinen Arbeitsbereich geeignet erscheinende Systematik ausgedacht hat, die nur in seinem Kopf existiert und im Laufe der Zeit durch neue Erfordernisse und/oder die den Menschen eigene Inkonsequenz modifiziert worden ist. Denkbare Kriterien sind neben der Chronologie und dem Alphabet der Korrespondenzpartner Leitzeichen – seltener Geschäfts- oder Aktenzeichen –, Sachbetreffe, kreative Mischformen und nicht minder originelle Abkürzungen. Eine für Außenstehende nachvollziehbare Dokumentation einer solchen Ablagesystematik existiert erfahrungsgemäß nicht.

Aber es geht noch weiter: Dieser Ansammlung von E-Mails ist nicht zu entnehmen, wer sie gesehen hat, oft noch nicht einmal, wer sie geschrieben hat oder an wen sie gerichtet sind. Ein typisches Beispiel lässt das deutlich vor Augen treten: „Sehr geehrter Herr Müller, ich stimme Ihrem Vorschlag zu. Mfg gez. Dr. Schröder“. Das ist alles, weitere Angaben finden sich in dieser Mail, abgesehen vom Sende- bzw. Empfangszeitpunkt, nicht. Der Kopfzeile kann man noch entnehmen, dass ein Dr. Schröder, Max an einen Müller, Herbert geschrieben hat. Sollten darüber hinaus die E-Mail-Adressen der Beteiligten erkennbar sein, lassen sich gewisse Rückschlüsse auf ihren beruflichen Kontext treffen. Aber weder die Funktionseinheiten noch die Positionen der beiden Korrespondenzpartner sind ersichtlich. Die Betreffzeile ist oft nicht oder unpräzise ausgefüllt.

Ich breche hier ab. Jeder von uns kennt derartige E-Mails und kann sie nur aufgrund der ihm individuell vorliegenden Kontextinformationen deuten. Ohne diese Person jedoch verlieren diese E-Mails deutlich an Aussagekraft.

Forschung in der digitalen Welt, die Sicherung, Erschließung und Aufbereitung von Wissensbeständen ist nicht nur ein technisches und/oder finanzielles Problem. Insbesondere die historische Forschung, allen voran die Geschichtswissenschaft, muss sich fragen, wieweit digitale Quellen auch eine modifizierte Methodik erfordern. Dabei kann sie von archivischer Seite unterstützt werden. Den Archiven kommt nämlich beim Informationstransfer zwischen Verwaltung und Wissenschaft in der digitalen Welt eine zentrale Rolle zu. Durch die Veränderungen in den Registratur führenden Stellen, seien sie in Behörden, Firmen, Vereinigungen oder bei Privatpersonen, verändert sich die Arbeit der Archive – und auch ihre Nutzerinnen und Nutzer müssen diese Entwicklungen nachvollziehen, wenn sie

eines Tages die Quellen des späten 20. und des 21. Jahrhunderts auswerten wollen.¹

2. Globalisierung: Digitale Unterlagen in der Verwaltung

Unternehmen der Privatwirtschaft werden durch ihre Kunden, durch die Globalisierung, durch neue Technologien und weitergehende Qualitätsanforderungen genau so zum Handeln herausgefordert wie Verwaltungen im öffentlichen Dienst als Erbringer von Dienstleistungen für die Bürgerinnen und Bürger.² Ein zentrales Element der Verwaltungsmodernisierung in den öffentlichen Verwaltungen sind Electronic-Government-Konzepte.³ Die von der Bundesregierung im September 2000 gestartete Initiative *BundOnline2005* betont, dass eine Informationsgesellschaft nicht ohne E-Government denkbar ist. E-Government bezieht sich auf alle Aspekte von Regierung und Verwaltung, deren Arbeit durch den Einsatz von Informations- und Kommunikationstechnologien unterstützt wird. Die Ziele einer solchen Modernisierung sind eine effizientere und rationellere Verwaltung, ein erhöhter interner Informationsaustausch, größere Transparenz des Verwaltungshandelns und somit eine höhere Qualität der Dienstleistung.⁴ In diesem Zusammenhang müssen daher auch die Verwaltungsprozesse zu Beginn der Implementierung von Dokumentenmanagementsystemen analysiert und optimiert werden, damit sie in der Verwaltung eine hohe Akzeptanz finden und erfolgreich eingesetzt werden können.⁵ Die Schriftgutverwaltung ‚hinter den Kulissen‘ muss effektiv, medienbruchfrei und den rechtlichen Vor-

¹ In diesem Beitrag erfolgt eine Konzentration auf Sachakten.

² Michael Leistenschneider: Elektronische Signatur – Motor für E-Government. In: *Innovative Verwaltung* 4 (2004). S. 44 f. – Rainer Pitschas: Neues Verwaltungsrecht im partnerschaftlichen Rechtsstaat? Zum Wandel von Handlungsverantwortung und -formen der öffentlichen Verwaltung am Beispiel der Vorsorge für innere Sicherheit in Deutschland. In: *Die Öffentliche Verwaltung* 6 (2004). S. 231–238, S. 231 f.

³ Lars Lentfer: Die Notwendigkeit von E-Government-Strategien. In: *Innovative Verwaltung* 10 (2002) S. 27 f. Vor übertriebenen Erwartungen im Kontext von E-Government-Konzepten warnen: Olaf Winkel: Zukunftsperspektive Electronic Government. In: *Aus Politik und Zeitgeschichte* 18, 2004. S. 7–15.

⁴ Archive und ihre Nutzer – Archive als moderne Dienstleister. Beiträge des 8. Archivwissenschaftlichen Kolloquiums der Archivschule Marburg. Hg. von Stefanie Unger (Veröffentlichungen der Archivschule Marburg, Institut für Archivwissenschaft 39). Marburg 2004.

⁵ Ralf Heib: Nur der Nutzen zählt. In: *Move Moderne Verwaltung* November 2003. S. 22–25.

gaben entsprechend funktionieren, damit der Verkehr nach außen, die Kommunikation mit den Bürgerinnen und Bürgern, bestmöglich gestaltet werden kann.

Verbunden damit ist, dass die Einführung dieser neuen Technologien auch mental begleitet wird, damit sie optimal eingesetzt werden können. Die neue Technik darf also nicht äußerlich aufgepfropft, sie muss innerlich angeeignet werden. Das Bewusstsein für einen ordnungsgemäßen Umgang mit den elektronisch gespeicherten Unterlagen ist erforderlich. Hier sind unterstützend die Archivarinnen und Archivare gefordert, die Behörden und Ämter gerade bei der Einführung von Dokumentenmanagementsystemen so zu begleiten, dass eine geordnete Schriftgutverwaltung gesichert ist.⁶

Dieser Aneignungsprozess vollzieht sich erfahrungsgemäß nicht automatisch und nicht von heute auf morgen. Auch wenn Archivarinnen und Archivare die Modernisierung sachkundig begleiten, ist leider davon auszugehen, dass in den Verwaltungen – und dies gilt auch für den privat(rechtlich)en Bereich – die Schriftgutverwaltung auf absehbare Zeit nicht optimal funktioniert, dass man also aussagekräftige Unterlagen nicht in der Ordnung findet, wie man sie erwarten könnte. Das zu wissen, ist für die historische Forschung von zentraler Bedeutung. Die Recherchestrategien müssen diesem Defizit angepasst werden, und zugleich müssen bei der Interpretation der ermittelten Unterlagen mögliche Überlieferungslücken einbezogen werden.

3. Digitale Unterlagen in den Archiven

Für die Archive bedeutet die Umstellung auf ausschließlich elektronische Speicherung von Unterlagen, dass sich ihre Arbeit immer mehr auf das

⁶ Margit Ksoll-Marcon: Digitale Unterlagen – eine neue Herausforderung bei der Behördenberatung. In: Archive und ihre Nutzer – Archive als moderne Dienstleister. Beiträge des 8. Archivwissenschaftlichen Kolloquiums der Archivschule Marburg. Hg. von Stefanie Unger (Veröffentlichungen der Archivschule Marburg, Institut für Archivwissenschaft 39). Marburg 2004. S. 225–237. – Rainer Hering: Die Mentalitätsoffensive als zentrales Instrument der Archivierung von Unterlagen aus digitalen Systemen. In: *Scrinium* 58 (2004). S. 80–87. – Digitales Verwalten – Digitales Archivieren. 8. Tagung des Arbeitskreises *Archivierung von Unterlagen aus digitalen Systemen* am 27. und 28. April 2004 im Staatsarchiv Hamburg. Hg. von Rainer Hering und Udo Schäfer (Veröffentlichungen aus dem Staatsarchiv der Freien und Hansestadt Hamburg 19). Hamburg 2004.

Vorfeld, das heißt die Schriftgut produzierenden und die registraturführenden Stellen konzentrieren muss. Die archivischen Kontakte zu den Verwaltungen im Sprengel müssen so ausgebaut und intensiviert werden, dass die Archivarinnen und Archivare rechtzeitig von der Entwicklung entsprechender Verfahren Kenntnis erlangen und an wichtiger Stelle – zum Beispiel in begleitenden Gremien – beteiligt werden können. Die Qualität der Schriftgutverwaltung in den Behörden und Ämtern muss geprüft werden, damit diese bei den Implementierungsprozessen in den Verwaltungen adäquat beraten werden können.

Diese Einbeziehung in das Verwaltungshandeln ist für die Archive von zentraler Bedeutung, da sich bei ausschließlich in digitalen Systemen gespeicherten Unterlagen die Bewertungsentscheidung von der Schließung einer Akte auf ihre Eröffnung vorverlagern wird. Sobald ein Aktentitel gebildet worden ist, ist also bereits eine Entscheidung über die Archivwürdigkeit zu treffen. Daher ist die richtige Zuordnung und Erfassung der einzelnen Schriftstücke unbedingt erforderlich. Die archivische Erschließung wird sich in der Regel auf die Übernahme der Metadaten elektronischer Unterlagen konzentrieren. Dadurch entfällt ein großer Teil der bisherigen archivischen Tätigkeit, doch setzt eine solche Umstellung intensive Vorfeldarbeit voraus, damit die dann als archivwürdig übernommenen Unterlagen auch benutzt werden können. Eine Sichtung der einzelnen Akten oder gar ein Neu- oder Umsortieren von falsch abgelegten Dokumenten wird in der Regel nicht mehr erfolgen können.

Zusammenfassend lässt sich also festhalten, dass sich durch die Umstellung auf die ausschließliche elektronische Speicherung von Verwaltungsunterlagen das Schwergewicht des archivischen Handelns von der retrospektiven auf die prospektive Arbeit verlagert. Diese Veränderung und ihre Konsequenzen sollten die Historikerinnen und Historiker bei der Auswertung archivischer Quellen im Blick haben.

4. Digitale Unterlagen und Historische Hilfswissenschaften

Die künftig in Archiven überlieferten Quellen der historischen Wissenschaften und ihre Entstehungsprozesse ändern sich also. Daher muss die Frage gestellt werden, ob nicht auch entsprechend die Methodik der Geschichtswissenschaft, auf jeden Fall die Quellenkunde, weiter entwickelt

werden muss. Das bisherige Angebot der Historischen Hilfs- bzw. Grundwissenschaften umfasst unter anderem Paläographie, Diplomatik, Sphragistik, Heraldik, Genealogie, Numismatik und Geldgeschichte, Historische Geographie, Urkunden- und Aktenlehre.⁷ Die relativ junge historische Bildkunde sorgt dafür, dass im visuellen Zeitalter Bilder und Fotos nicht nur als mehr oder weniger zu den Texten passende Illustrationen verwendet, sondern auch zunehmend als historische Quellen methodisch fundiert analysiert werden.⁸ Paläographie und Aktenkunde bieten notwendige Informationen und Fähigkeiten zum Verständnis von Schriftgut, insbesondere Urkunden und neuzeitlicher Akten. Anredeformen, Aufbau von Schriftstücken, Vermerke, Geschäftszeichen etc. sind so interpretierbar – auch wenn man manchmal den Eindruck gewinnt, dass nicht immer in der erforderlichen Weise über den Text hinaus auch der Kontext eines Dokuments wahrgenommen wird.

Wie aber sollen angesichts der eingangs geschilderten Situation nur noch elektronisch überlieferte Unterlagen angemessen gesichtet und interpretiert werden? Die vielfach erträumte vollständige Überlieferung aller entstandenen Dokumente ist langfristig durch die erforderliche Migration der Daten wohl nicht zu bewältigen und nicht sinnvoll. Eine denkbare Volltextrecherche wird aufgrund ihrer eingeschränkten Aussagekraft – hat man wirklich den richtigen Begriff eingegeben? – und der zu erwartenden exorbitanten Trefferzahl nicht weiter helfen. Zudem: Die Gefahr, ein Schriftstück vollends aus seinem Entstehungszusammenhang zu isolieren und es somit nicht angemessen interpretieren zu können, ist groß. Der Kontext bildet auch im digitalen Zeitalter den Rahmen für seine Deutung. Es ist eben nicht egal, in welchem

⁷ Ahasver von Brandt: *Werkzeug des Historikers. Eine Einführung in die historischen Hilfswissenschaften*. 9., erg. Aufl. Stuttgart, Berlin, Köln, Mainz 1980. – Heinrich Otto Meisner: *Archivalienkunde vom 16. Jahrhundert bis 1918*. Leipzig 1969. – *Die archivalischen Quellen. Mit einer Einführung in die Historischen Hilfswissenschaften*. Hg. von Friedrich Beck, Eckart Henning. 3., überarb. u. erw. Aufl. Köln, Weimar, Wien 2003. – Hans Wilhelm Eckardt, Gabriele Stüber und Thomas Trumpp unter Mitarbeit von Andreas Kuhn: *Paläographie – Aktenkunde – Archivalische Textsorten. „Thun kund und zu wissen jedermänniglich“* (Historische Hilfswissenschaften bei Degener & Co. 1). Neustadt/Aisch 2005. – *Historische Hilfswissenschaften. Stand und Perspektiven der Forschung*. Hg. von Toni Diederich und Joachim Oepen. Köln, Weimar, Wien 2005.

⁸ Rainer Wohlfeil: *Das Bild als Geschichtsquelle*. In: *Historische Zeitschrift* 243 (1986). S. 91–100. – *Historische Bildkunde. Probleme – Wege – Beispiele*. Hg. von Brigitte Tolkemitt und Rainer Wohlfeil (*Zeitschrift für Historische Forschung*, Beiheft 12). Berlin 1991. – Heike Talkenberger: *Von der Illustration zur Interpretation: Das Bild als historische Quelle. Methodische Überlegungen zur Historischen Bildkunde*. In: *Zeitschrift für Historische Forschung* 21 (1994). S. 289–313.

Zusammenhang ein Schriftstück entstanden oder überliefert worden ist. Seine Zuordnung zu einer Akte ist ebenfalls interpretationsrelevant.

Gefragt werden muss dabei nicht nur nach dem, was überliefert ist, sondern gerade auch nach dem, was nicht vorhanden ist – und vor allem aus welchen Gründen. Sind die erwarteten Unterlagen gar nicht entstanden? Sind sie nicht oder nicht am erwarteten Ort gespeichert worden? Sind sie nicht dem zuständigen Archiv angeboten worden? Oder hat dieses sie nicht für archivwürdig befunden?

Zudem: Im Bereich der Neueren und Neuesten Geschichte wird vergleichsweise selten nach der Echtheit eines Schriftstückes gefragt. Die angeblichen Hitler-Tagebücher stellten eine herausragende, intensiv in der Öffentlichkeit diskutierte Ausnahme dar. Eine Akte in einem Archiv reicht zumeist als Garantie für eine ungebrochene und unveränderte Überlieferung aus. Wie aber ist das im digitalen Zeitalter? Wie leicht elektronisch gespeicherte Unterlagen verändert werden können, ist nur zu bekannt. Intensiv wird an digitalen Signaturen und ihren Einsatzmöglichkeiten gearbeitet.⁹

Die Liste von Fragen ließe sich unschwer verlängern. Deutlich wird, dass eine methodische Reflexion erforderlich ist, um die Auswirkungen, die die digitale Welt auf der Ebene der Quellen für die historische Forschung hat, zu untersuchen. Ist eine Weiterentwicklung der Historischen Hilfswissenschaften erforderlich?

Auf einem Workshop der Deutschen Forschungsgemeinschaft in Münster im Oktober 2004 über die Perspektiven der Kooperation zwischen Geschichtswissenschaft und Archiven wurde angesichts der fortschreitenden Reduzierung der Historischen Hilfswissenschaften an den Universitäten und der zu geringen hilfswissenschaftlichen Vorkenntnisse der Archivbesucher diskutiert, ob hier die Archive oder die Universitäten gefordert sind, diese Defizite abzubauen.¹⁰ Angesichts der finanziellen und personellen Lage der Archive,

⁹ Udo Schäfer: Elektronische Signaturen oder Ius Archivi? In: Digitales Verwalten – Digitales Archivieren. 8. Tagung des Arbeitskreises *Archivierung von Unterlagen aus digitalen Systemen* am 27. und 28. April 2004 im Staatsarchiv Hamburg. Hg. von Rainer Hering und Udo Schäfer (Veröffentlichungen aus dem Staatsarchiv der Freien und Hansestadt Hamburg 19). Hamburg 2004. S. 13–31. – Stefanie Fischer-Dieskau: Elektronisch signierte Dokumente. Anforderungen und Maßnahmen für ihren dauerhaften Erhalt. In: Ebd. S. 33–50. – Wolfgang Farnbacher: Technische und organisatorische Konzepte des ArchiSig-Projekts. In: Ebd. S. 51–65. – Jutta Drühmel: Digitale Signatur in der Praxis. Elektronischer Rechtsverkehr am Finanzgericht Hamburg. In: Ebd. S. 67–68.

¹⁰ Ragna Boden, Christine Mayr, Christoph Schmidt und Thomas Schwabach: Tagungsbericht: Die Geschichtswissenschaften und die Archive. Perspektiven der Kooperation. DFG-Workshop

die zweifelsohne nicht günstiger als die der Universitäten ist, kann diese Aufgabe nicht von ihnen allein gelöst werden. Aber sie können gerade angesichts ihrer veränderten Rolle im Informationstransfer zwischen Verwaltung und Wissenschaft gemeinsam mit den Historischen Hilfswissenschaften die erforderliche Weiterentwicklung und Vermittlung des methodischen Rüstzeugs vorantreiben. Die Bedeutung der Historischen Hilfswissenschaften wächst angesichts der geschilderten Entwicklung. Wenn dies gesehen und konstruktiv genutzt wird, wird es eher möglich sein, die 10 000 E-Mails des einleitend zitierten leitenden Beamten einer Hamburger Behörde quellenkritisch und im Kontext für die historische Forschung nutzbar zu machen und insgesamt eine Erosion des kulturellen Gedächtnisses zu verhindern.¹¹

am 05.10.2004 in Münster (H-SOZ-U-KULT 29.10.2004).

¹¹ Manfred Osten: Digitalisierung und kulturelles Gedächtnis. In: *Aus Politik und Zeitgeschichte* 5–6/2006, S. 3–8.

Nutzung von Digitalisaten am Beispiel des Geheimen Staatsarchivs Preußischer Kulturbesitz

Dieter Heckmann

Bevor die Nutzung im eigentlichen Sinne zur Sprache kommt, sei es mir gestattet, zunächst den Beginn, den derzeitigen Stand und die mittelfristigen Perspektiven der Digitalisierung im Geheimen Staatsarchiv in wenigen Strichen zu skizzieren. Im Anschluss daran möchte ich auf die Nutzung der elektronisch erzeugten Bilder im Rahmen der Gebührenordnung zu sprechen kommen. Dabei sollen nicht nur Entgelte erläutert, sondern auch urheberrechtliche Fragen berührt werden. Zum Schluss gilt es, einige Vorstellungen und Gedanken über Nutzungsbeschränkungen und Nutzungswerte von Digitalisaten vorzustellen.

1.

Das Gutachten einer Wirtschaftsprüfungsgesellschaft, das der Präsident der Stiftung Preußischer Kulturbesitz für seine Einrichtungen 1993 in Auftrag gegeben hatte, schlug hinsichtlich der künftigen Archivierungs- und Sicherungsstrategie des Geheimen Staatsarchivs unter anderem die Digitalisierung vor. Ohne auf Einzelheiten eingehen zu wollen, sah das Gutachten bereits zwei Arten der Digitalisierung vor, nämlich einmal die unmittelbare Umsetzung der Vorlagen mittels direkter Abtastung und andermal die Scannung der massenhaft vorhandenen Mikrofilme. Bei der zuletzt genannten Art sollte nicht die Kopie des Sicherungsfilms, sondern die Kopie der Kopie, also die dritte Generation eines Films, zum Einsatz kommen. Hauptziele dieser Gutachervoten waren die Schonung der Bestände und die Steigerung der Effizienz. In Anbetracht der Benutzungshäufigkeit er-

scheinen diese Ziele nach wie vor geradezu geboten. Das Geheime Staatsarchiv gehört nämlich mit rund 11 745 Benutzertagen jährlich und durchschnittlich 4176 schriftlichen Anfragen im Jahr¹ zu den am häufigsten besuchten Archiven in der Bundesrepublik Deutschland. Die Digitalisierung von besonders nachgefragten Beständen oder Bestandteilen hatte deswegen Vorrang. Das Geheime Staatsarchiv ließ in den Jahren 1997 und 1998 die ersten Schutzfilme, eine Anzahl von Findhilfsmitteln und sonstiges ausgewähltes Schriftgut im Auftragswert von jeweils rund 350 000 DM digitalisieren.² Stichproben ergaben, dass die Wiedergabequalität jedoch viel zu wünschen übrig ließ. Dies konnte besonders bei den Aufnahmen handgeschriebener Vorlagen festgestellt werden. Allerdings hängt die schlechte Qualität der Schutzfilme ursächlich mit der der frühen Sicherungsfilme zusammen, die wiederum mit dem Beginn der Sicherungsverfilmung in den Staatsarchiven Ende der 60er Jahre des vorigen Jahrhunderts im Zusammenhang steht. Damals waren nämlich die heutigen Verfilmungsstandards erst in bescheidenen Ansätzen entwickelt. Die genannten Qualitätsprüfungen waren gerade für das Geheime Staatsarchiv besonders wichtig, weil die Masse der Bestände aus handgeschriebenem Schriftgut besteht, welches besonders gute bildliche Wiedergaben verlangt. Dass ungeachtet dessen die Digitalisierung der Rollfilmbestände auch nach 1999 fortgesetzt wurde, hing mit den mangelnden Alternativen zusammen. Seit 2002 ist allerdings die Digitalisierung der Schutzfilme wegen fehlender Haushaltsmittel ausgesetzt worden. Im vergangenen Jahr konnten dennoch 211 Meter Rollfilm mittels eines Dupliziergerätes in hauseigener Regie kopiert werden. Die Anzahl der bis 2005 hergestellten Images beträgt rund 4,5 Millionen. Diesen 4,5 Millionen Bildern liegen 2 450 Schutzfilme zugrunde.³ Benutzbar sind die Images jedoch noch nicht, weil die Verknüpfung mit den Metadaten fast ausnahmslos fehlt. Hier ist noch ein riesiger Arbeits- und Kostenaufwand erforderlich, bis die Koppelung an die in der *Augias-Datenbank* gespeicherten Findbucheinträge bewerkstelligt ist. Erste Ergebnisse sollen aber im Laufe diesen Jahres für die Online-Recherche im Forschungsaal des Geheimen Staatsarchivs zur Verfügung stehen. Auf eine preußische

¹ Zugrunde liegen die aus den Jahresberichten ermittelten Durchschnitte der Jahre von 1996 bis 2005.

² Jahresbericht 1997, S. 34.

³ Laut Auskunft des zuständigen Referenten vom 9. März 2006 seien rund 4,5 Millionen Images von 2450 Schutzfilmen auf circa 6800 Compact Disks zu 4 Terabytes gebrannt worden.

Akte von durchschnittlich 250 Blatt kämen so gerechnet 500 Images, was bei 4,5 Millionen digitaler Bilder 9 000 Akteneinheiten oder 1 000 laufende Meter Archivgut bedeutet. Trotz dieser gewaltigen Menge an digitalen Aufnahmen ist somit erst ein Bruchteil des Gesamtumfanges digitalisiert. Der Rest von rund 34 laufenden Kilometern Schriftgut wäre noch entsprechend zu bearbeiten und vor allem zu finanzieren. Vielleicht bietet die kurz vor der industriellen Einführung stehende Hologrammtechnik in naher Zukunft soviel an Anreizen, dass damit die heute bekannten Speicherträger für Digitalisate kostengünstiger und wirkungsvoller ersetzt werden können.⁴

Dem angesprochenen Digitalisierungsaufwand hat das Geheime Staatsarchiv auch eine herkömmliche Alternative – nicht zuletzt dank der noch vorhandenen paläographischen Lesefähigkeiten – entgegensetzen können, nämlich die Bestandserschließung mittels Vollregesten. Im gleichen Zeitraum, also von 1997 bis 2005, haben eine halbe Projektstellenkraft und zeitweilig zwei Referatsleiter des Geheimen Staatsarchivs immerhin rund 10,5 laufende Meter frühneuzeitliches Schriftgut regestiert und in Buchform der Öffentlichkeit zur Verfügung gestellt.⁵ Während sich die Wertschöpfung bei der Schutzfilmdigitalisierung noch erweisen muss, liegt der volkswirtschaftliche Nutzen bei den veröffentlichten Vollregesten klar auf der Hand:

⁴ Siehe dazu: Cornelia Denz: Volumenhologramme – Datenspeicher der Zukunft. In: *Physikalische Blätter* 55, 1999 Nr. 4. S. 41–45; Universität Münster entwickelt holographischen Speicher. In: *Golem.de*. Adresse: <http://www.golem.de/0403/30558.html> (letzte Einsichtnahme am 22.03.2006).

⁵ Dieter Heckmann (Bearb.): *Die Beziehungen der Herzöge in Preußen zu West- und Südeuropa (1525–1688)*. Regesten aus dem Herzoglichen Briefarchiv und den Ostpreußischen Folianten (Veröffentlichungen aus den Archiven Preußischer Kulturbesitz 47). Köln, Weimar, Wien 1999; Stefan Hartmann (Bearb.): *Herzog Albrecht von Preußen und Livland (1534–1540)*. Regesten aus dem Herzoglichen Briefarchiv und den Ostpreußischen Folianten (Veröffentlichungen aus den Archiven Preußischer Kulturbesitz 49). Köln, Weimar, Wien 1999; Stefan Hartmann (Bearb.): *Herzog Albrecht von Preußen und Livland (1540–1551)*. Regesten aus dem Herzoglichen Briefarchiv und den Ostpreußischen Folianten (Veröffentlichungen aus den Archiven Preußischer Kulturbesitz 54). Köln, Weimar, Wien 2002; Stefan Hartmann (Bearb.): *Herzog Albrecht von Preußen und Livland (1551–1557)*. Regesten aus dem Herzoglichen Briefarchiv und den Ostpreußischen Folianten (Veröffentlichungen aus den Archiven Preußischer Kulturbesitz 57). Köln, Weimar, Wien 2005; Ursula Benninghoven (Bearbeiterin): *Die Beziehungen Herzog Albrechts von Preußen zu Städten, Bürgertum und Adel im westlichen Preußen (1525–1554)*. Regesten aus dem Herzoglichen Briefarchiv und den Ostpreußischen Folianten (Veröffentlichungen aus den Archiven Preußischer Kulturbesitz 48, 1–2). Köln, Weimar, Wien 2006; Stefan Hartmann (Bearb.): *Herzog Albrecht von Preußen und Livland (1557–1560)*. Regesten aus dem Herzoglichen Briefarchiv und den Ostpreußischen Folianten (Veröffentlichungen aus den Archiven Preußischer Kulturbesitz 60). Köln, Weimar, Wien 2006.

Der Leser kann die mittels der Regesten aufbereiteten und im Buch zur Verfügung gestellten Nachrichten bequem ohne besondere Vorkenntnisse sprachlicher, paläographischer oder sonstiger Art und ohne zeit- und kostenträchtige Archivbesuche für seine Zwecke verwenden.

Der Vollständigkeit halber sei in diesem Zusammenhang in wenigen Worten noch der Anteil der Direktscans erwähnt. Er beträgt derzeit nur wenige hundert Bilder. Bei diesen handelt es sich in der Regel um die Ablichtung ausgewählter Stücke, zumeist aus der Kartenabteilung, die für Kataloge, Prospekte und dergleichen hergestellt wurden. Trotz ihrer geringen Anzahl verblüffen die Images durch ihre gute Qualität. Sie ist vor allem dem Einsatz moderner Aufnahme- und Wiedergabetechnik geschuldet.

An direkten Aufnahmen sind in zunehmendem Maße Archivbenutzerinnen und Archivbenutzer interessiert, die mit Hilfe eigener Digitalkameras Ablichtungen vornehmlich aus Akten anzufertigen wünschen. Diesem Ansinnen hat die Benutzungsordnung des Geheimen Staatsarchivs einen Riegel vorgeschoben. In der seit dem 2. Januar diesen Jahres in Kraft befindlichen Ordnung⁶ heißt es nämlich unter § 8 Absatz 1: *Die Anfertigung von Reproduktionen durch Benutzerinnen bzw. Benutzer ist grundsätzlich nicht zulässig. Ausnahmen bedürfen der ausdrücklichen Genehmigung durch den Direktor.* Diese Regelung bedeutet keine Neuerung, denn sie findet sich wörtlich in der alten Benutzungsordnung vom 9. Juni 1995 wieder. Stand hinter der Formulierung von 1995 aber noch die Sorge um den Bestandserhalt im Vordergrund, so ist nun diese (im Hinblick auf die die Vorlagen schonende Abtasttechnik) dem Gesichtspunkt der Verwertung gewichen. Angesichts knapper werdender Haushaltsmittel ist das Archiv mehr denn je auf Einnahmen von Dienstleistungen aus der hauseigenen Bildstelle angewiesen. So kostet beispielsweise eine digitale Farbaufnahme 18 Euro. Derselbe Preis gilt auch für einen Color-Durchlicht-Scan. Der Auftraggeber hat sogar 32 Euro zu entrichten, wenn er eine digitale Aufnahme von einer dreidimensionalen Vorlage, wie zum Beispiel von einem Aktenband, von einer Urkunde, einem Siegel oder von einer Medaille, wünscht. Die Höhe dieser Gebühren soll einen Mittelwert zwischen den reinen Herstellungskosten und den marktüblichen Preisen widerspiegeln.

Die Nutzung der bei der Bildstelle des Geheimen Staatsarchivs in Auftrag gegebenen Aufnahmen ist an strenge Auflagen geknüpft. So wird die

⁶ Adresse: <http://www.gsta.spk-berlin.de/framesets/frameset.php> (letzte Einsichtnahme am 27.03.2006).

Auftraggeberin oder der Auftraggeber eigens auf folgende Bedingungen hingewiesen:

„Die von der Bildstelle des Geheimen Staatsarchivs PK angefertigten Reproduktionen sind nur zum direkten persönlichen Gebrauch der Auftraggebenden bestimmt. Alle Nutzungs- und Verwertungsrechte an den Reproduktionen liegen beim Geheimen Staatsarchiv PK. Die Weitergabe oder Überlassung der Reproduktionen an Dritte sowie die Vervielfältigung in jedweder Art (Reproduzieren, Kopieren, Digitalisieren, Duplizieren, Archivieren, Scannen, Speichern etc.) zum Zweck der Weitergabe oder Überlassung an Dritte sind verboten. Digitale Reproduktionen sind ohne Aufforderung unmittelbar nach dem persönlichen Gebrauch zu löschen. Jede Nutzung über den persönlichen Gebrauch hinaus, insbesondere die Veröffentlichung, bedarf der Genehmigung des Geheimen Staatsarchivs PK. Eine Veröffentlichung ist honorarpflichtig, soweit sie gewerblichen Zwecken dient. Diese Bedingungen gelten für alle vom Geheimen Staatsarchiv PK zur Verfügung gestellten Reproduktionen. Sie gelten auch dann, wenn das Bildmaterial über Dritte oder aus anderen Quellen in Besitz genommen wird (vergleiche Benutzungsordnung § 8 Ziff. 3 i. d. F. vom 9. Dezember 2005).“⁷

2.

Stellt eine Benutzerin oder ein Benutzer trotz des Verbotes in der Benutzungsordnung eigenmächtig Aufnahmen von Archivalien des Geheimen Staatsarchivs her, droht ihr oder ihm höchstens ein zeitlich begrenztes Hausverbot; hält sich doch der Schaden in Grenzen, der mit einem derartigen Verhalten verursacht wird. Selbst bei heimlich angefertigten Aufnahmen mittels einer eigenen Digitalkamera erwirbt – wie oben bereits dargelegt – die Benutzerin beziehungsweise der Benutzer weder Verwertungs- noch Urheberrechte. Davon abgesehen gilt die Anfertigung einer Digitalaufnahme nicht im Sinne von § 2 Absatz 2 des Urhebergesetzes als geschütztes Werk, denn der eigenschöpferische oder gestalterische Anteil bei der Erzeugung eines

⁷ Aus der Sammlung der allgemeinen Textbausteine entnommen (letzte Einsichtnahme am 27.03.2006).

solchen Bildes ist nicht oder kaum erkennbar. In dieser Hinsicht lässt sich die Digitalaufnahme mit einer Fotokopie vergleichen.

Die angesprochenen Dienstleistungen der Bildstelle schließen auch die Nutzung der Urheberrechte durch das Archiv ein. Dies gilt in der Regel uneingeschränkt für diejenigen Photographinnen oder Photographen, die mit dem Archiv arbeitsrechtlich durch unbefristete Verträge verbunden sind, denn die Überlassung der Nutzungsrechte an den eigenen Bildern ist üblicherweise Bestandteil des Anstellungsvertrages mit dem Arbeitgeber. Bei zeitlich befristeten Arbeitsverhältnissen oder gar bei Werkverträgen sieht die Rechtslage vielfach anders aus. Gerade bei Werkverträgen sollte die Arbeitgeberin oder der Arbeitgeber tunlichst darauf achten, dass ihr oder ihm die ausschließlichen Nutzungsrechte vertraglich überlassen werden.⁸ Sonst haben die entsprechende Photographin oder der Photograph oder die Erben das Recht, die im Rahmen des Vertrages entstandenen Aufnahmen selbst zu verwerten, was zum Beispiel bei häufig nachgefragten Bildern aus Katalogen zu erheblichen Einnahmen für sie oder ihn oder umgekehrt zu großen Einnahmeverlusten für das Archiv führen kann.

Bekanntermaßen erlöschen die Urheberrechte einer Photographin oder eines Photographen oder ihrer oder seiner Erben erst 70 Jahre nach dem Todesfall. Dank dieser Regelung kann das Geheime Staatsarchiv die Sammlung seiner Digitalisate, die nach § 2 Absatz 2 des Urhebergesetzes als Werk geschützt sind, zügig erweitern, um eine künftige Einnahmequelle einzufassen.

3.

Was die weiteren Beschränkungen in der Nutzung einerseits sowie andererseits die Nutzungswerte von Digitalisaten anbelangt, so heißt es zwar im § 4 Absatz 1 der Benutzungsordnung des Geheimen Staatsarchivs, dass das Archivgut und dessen hand- oder schreibmaschinenschriftliche Findmittel sowie das Bibliotheksgut der Dienstbibliothek nur im Forschungssaal des Archivs benutzt werden. Diese Regelung wird aber in dem Maße durchlöchert, je mehr Findmittel in digitaler Form online oder auch in gedruckter Form angeboten werden. Beispiele dafür geben sowohl die schon online re-

⁸ Jan A. Strunk: Die Geister, die ich rief ... In: Computer-Fachwissen 1 (2004). S. 27–30, S. 30.

cherchierbaren Karten als auch die gedruckten Regestenwerke, die die vollständige Signatur jedes einzelnen Stücks mitteilen, das registriert worden ist. Diese Schwachstelle in der Benutzungsordnung lässt sich jedoch leicht umkehren, wenn mit der Erlaubnis zur aushäusigen Benutzung, zum Beispiel via Internet, künftig Entgelte verbunden werden. Entsprechende Vorbilder aus einigen Bundesländern⁹ liegen bereits vor.

Wie bereits angedeutet, sollen im Verlauf dieses Jahres die ersten Verknüpfungen zwischen Metadaten und den dazu gehörenden Images für die Bildschirmrecherche im Benutzersaal des Geheimen Staatsarchivs zur Verfügung gestellt werden. Mit diesem Pilotprojekt sollen die Nachrichten aus denjenigen Akten wieder zugänglich gemacht werden, die wegen Entsäuerungsmaßnahmen¹⁰ für die Benutzung unzugänglich geworden sind. Wie früher die Mikrofiches sollen die Images heute als Ersatz für die gesperrten Akten dienen. Da diese Akten für die Restaurierung aus ihren Heftungen gelöst und jetzt mittels ‚Chemises‘ – darunter sind Umschläge mit Seitenklappen zu verstehen – zusammengehalten werden, sind die losen Akten nur noch aus triftigen Gründen vorlegbar. Ein triftiger Grund wäre etwa gegeben, wenn versehentlich eine Seite nicht verfilmt ist. Verlangt jedoch der Benutzer wegen mangelnder paläographischer Lesefähigkeit die Herausgabe der Papiervorlage, so wird ihm in der Regel die Einsicht verwehrt.

Sollten die Haushalte jedoch – und das ist zu befürchten – mittelfristig eine Verknüpfung der bereits jetzt erzeugten 4,5 Millionen Images des Geheimen Staatsarchivs mit ihren Metadaten nicht mehr zulassen, dürfte der soeben beschriebene Einsatz von Digitalisaten zu den vernünftigsten Arten der Umsetzung gehören. Von daher sollte der Verwendung von Digitalisaten zum Schutz der vom Zerfall bedrohten Vorlagen unbedingt Vorrang eingeräumt werden.

⁹ Sowohl in Bremen (über Staatsarchiv Bremen, Präsident-Kennedy-Platz 2, 28203 Bremen) als auch in Hamburg wurden seit den 1830er Jahren von den Schiffsmaklern Verzeichnisse mit persönlichen Daten und Angaben zur Herkunft der Passagiere für die Polizeibehörden zusammengestellt (Adresse: <http://linktoyourroots.hamburg.de>, letzte Einsichtnahme am 26.04.2006). Eine Namensrecherche ist nur gegen Gebühr beim Staatsarchiv der Freien und Hansestadt Hamburg, Kattunbleiche 19, 22041 Hamburg, möglich.

¹⁰ Dazu kritisch Dieter Heckmann: Paläographie: eine archivistische Dienstleistung mit Zukunft. In: Regionen Europas – Europa der Regionen. Festschrift für Kurt-Ulrich Jäschke zum 65. Geburtstag. Hg. von Peter Thorau, Sabine Penth und Rüdiger Fuchs. Köln, Weimar, Wien 2003. S. 287–295, S. 292 f. Netzversion Adresse: <http://people.freenet.de/heckmann.werder/Palaographie.htm>.

Eine weitere Mengendigitalisierung scheint darüber hinaus derzeit nur noch bei ausgewählten Vorlagen finanzierbar zu sein. Die Auswahl sollte sich hierbei freilich von der Entwicklung der Nachfrage leiten lassen. So ist schon seit Jahren erkennbar, dass maschinengeschriebene Vorlagen weit oben in der Gunst der Benutzerinnen und Benutzer liegen und sich zudem wachsender Beliebtheit auch bei Wissenschaftlerinnen und Wissenschaftlern erfreuen. Das Lesen dieser Vorlagen bedarf nämlich keiner paläographischen oder sonstigen hilfswissenschaftlichen Vorbildung. Hinzu kommt, dass wegen der weitgehenden Normierung der maschinenschriftlichen Zeichen selbst Digitalisate von Aufnahmen mittelmäßiger oder gar schlechter Wiedergabequalität von den Benutzerinnen und Benutzern vorbehaltlos angenommen werden.

Das Angebot der Archive in der digitalen Welt

Retrokonversion, Datenaustausch und Archivportale

Frank M. Bischoff und Udo Schäfer

In seinem Beitrag über *Archive und Internet*¹ berichtete Karsten Uhde, dass im Jahr zuvor gerade einmal fünf archivische Einrichtungen in Deutschland – darunter zwei Ausbildungsstätten – mit einer Website im Internet vertreten waren. Die Veröffentlichung dieser Aussage liegt jetzt zehn Jahre zurück und bezog sich auf das Jahr 1995. Ein Blick auf die Linkliste der Archivschule Marburg² und insbesondere in die dort benannten Portale zeigt, dass sich dieser Zustand nachhaltig geändert hat. Es gibt keinen Bereich innerhalb des deutschen Archivwesens, der in der jüngeren Vergangenheit eine vergleichbar rasante Entwicklung erlebt hätte.

Das hat gute Gründe. Wollen Archive als Dienstleister gegenüber anderen Informations Providern konkurrenzfähig bleiben, müssen sie ihre Kunden verstärkt mit Online-Informationen im Internet versorgen. Das Werben um Benutzer und Kunden beginnt heute nicht mehr erst in den Lesesälen, sondern bereits im Internet. In einer Gesellschaft, die sich zur Informationsbeschaffung mit weiterhin zunehmender Tendenz der Recherchemöglichkeiten in Rechnernetzwerken bedient, gehen Einrichtungen, die dort nicht präsent sind, inzwischen das sehr reale Risiko ein, gar nicht wahrgenommen zu werden. Für den Internetbenutzer existieren sie de facto nicht.

Gerade deshalb müssen sich Archivare heute in dem Maße, in dem die eingehende Erforschung der eigenen Bestände im Alltag der Archivarbeit immer seltener ihren Platz finden kann, verstärkt als Informationsbroker

¹ Karsten Uhde: *Archive und Internet*. In: *Der Archivar* 49 (1996). Sp. 205–216, bes. 206.

² Vgl. Adresse: <http://www.archivschule.de/content/33.html> (letzte Einsichtnahme am 26.05.2006).

betätigen und ihre Bestände einem interessierten Publikum vermitteln. Was nachgefragt ist, sind Inhalte, also Beständeübersichten und Findbücher, die in weitaus größerem Umfang im Internet bereitgestellt werden müssen, als das heute der Fall ist. Bei allem Respekt vor der inzwischen beachtlichen Präsenz deutscher Archive im Internet bestehen auf diesem Gebiet noch erhebliche Desiderate. Ähnlich wie große Universitätsbibliotheken in Deutschland fast flächendeckend ihre Kataloge online zugänglich gemacht haben, sollten auch Archive ihre Findmittel – natürlich unter Beachtung persönlichkeitschutzrechtlicher Bestimmungen und gegebenenfalls anderer rechtlicher Einschränkungen – ins Netz stellen. Es handelt sich hier zugleich um eine ideelle Pflicht von Archiven im Rahmen ihres Charakters als Häuser der Geschichte, Gedächtnis der Gesellschaft und vor dem Hintergrund ihrer Mitverantwortung für die historisch-politische Bildung. Eine Digitalisierung von Archivgut, die Erstellung von Online-Zimelienschauen oder Internet-Ausstellungen gehören demgegenüber immer noch eher zur Kategorie der Kür. Damit soll keineswegs in Abrede gestellt werden, dass die letztgenannten Bereiche das Publikum ansprechen und werbewirksam sein können.

Im Hinblick auf die aktuellen Anforderungen an und Perspektiven für archivische Internet-Informationsangebote will dieser Beitrag im Folgenden drei Fragen nachgehen:

- Welche Bedeutung kommt Archivportalen zu?
- Welchen Stellenwert haben Schnittstellen und welche Ansprüche sollten sie erfüllen?
- Wie ist die Bereitstellung einer großen Zahl von Findmitteln im Internet realisierbar?

1. Archivportale

In 2005 verabschiedete der Rat der Europäischen Union eine Reihe von Empfehlungen über vorrangige Aktionen zur Stärkung der Zusammenarbeit im europäischen Archivwesen.³ Es handelt sich dabei um Maßnahmen zur Schadensprävention, um verstärkte interdisziplinäre Zusammenarbeit

³ Vgl. die deutsche Zusammenfassung unter Adresse: http://eur-lex.europa.eu/LexUriServ/site/de/oj/2005/l_312/l_31220051129de00550056.pdf (letzte Einsichtnahme am 26.05.2006).

bei elektronischen Unterlagen, um rechtliche Regelungen zu Verwaltung von und Zugang zu Schrift- und Archivgut sowie um vorbeugende Maßnahmen gegen Archivgutdiebstahl.

Neben diese vier Empfehlungen von eher generellem Charakter tritt eine fünfte, die sich im Vergleich zu den übrigen fast als Spezialauftrag liest. Der Rat der Europäischen Union empfiehlt nämlich, in den nächsten Jahren folgende Maßnahme gemeinsam durch Mitgliedstaaten und EU-Organen umzusetzen:

„Einrichtung und Betreuung eines Internet-Portals für Dokumente und Archive in Europa.“

Die Empfehlung wird wie folgt spezifiziert:

„Bereitstellung eines Internet-Portals durch die Nationalarchive der Mitgliedstaaten und die Archivdienste der EU-Organen, um den grenzüberschreitenden Zugang zu Dokumenten und Archiven der Mitgliedstaaten und der EU-Organen zu erleichtern. Dieses Internet-Portal könnte vom Nationalarchiv eines Mitgliedstaats betreut werden.“⁴

Der Empfehlung des Rats liegt der Bericht über die Archive in der erweiterten Europäischen Union zugrunde.⁵ Im Auftrag der Europäischen Kommission war dieser Bericht von einer Gruppe nationaler Sachverständiger mit dem Ziel eines Aktionsplans für eine erweiterte europäische Kooperation auf dem Gebiet des Archivwesens erarbeitet worden.

Der Sachverständigenbericht befasst sich in zwei Abschnitten⁶ mit Onlinezugang zu Findmitteln und Internetportalen. Die geforderte Einrichtung eines europäischen Internet-Gateways soll die bislang disparaten europäischen Internet-Initiativen bündeln und Benutzern einen One-Stop-Shop für den Zugang zu archivischen Informationen bieten. Erhöhte Benutzer-

⁴ Ebd., Ziff. B. 3.

⁵ Report on Archives in the enlarged European Union. Increased archival cooperation in Europe: action plan, elaborated by the National Experts Group on Archives of the EU-Member States and EU-Institutions and Organs, Adresse: http://www.archives.gov.ua/News/EuroUnion/report_Archives.pdf (letzte Einsichtnahme am 26.05.2006).

⁶ Ebd., Kapitel 2c und 2d, S. 55 ff.

freundlichkeit und die Bereitstellung einer kritischen Masse von archivischen Informationen auf europäischer Ebene ist das Ziel der aufgestellten Forderungen. Das europäische Archivportal soll nach Auffassung der nationalen Sachverständigen mindestens drei Schichten umfassen:

- Die Integration aller Daten aus archivischen Webpräsentationen in recherchierbarer Form.
- Eine Crawler-Abfrage auf Findmittel, die mangels geeigneter Schnittstellen nicht integrierbar sind, aber mit ‚Google-Techniken‘ recherchiert werden können.
- Eine Link-Liste zu bibliothekarischen Verbundkatalogen, wo gedruckte Findmittel nachgewiesen werden.

Allgemeine Informationen über die Archive, wie Adressen, Öffnungszeiten etc. sowie Hilfestellungen zu verschiedenen Themenbereichen sollen das Portal komplettieren.

Wie lässt sich diese ausgesprochen konkret formulierte Forderung nach einer Portallösung auf europäischer Ebene erklären? Dass die Europäische Union den Archiven ihrer Mitgliedsländer zu erhöhter Aufmerksamkeit und den Bürgern zu verbesserten Möglichkeiten der Informationsabfrage verhelfen möchte, scheint unmittelbar einleuchtend. Warum aber wird der Portalgedanke derart betont?

Um diese Forderung näher zu beleuchten, sei hier beispielhaft auf ein konkretes Portalprojekt zurückgegriffen. Als das nordrhein-westfälische Archivportal 1998 im Internet frei geschaltet wurde, war es das erste seiner Art. Es schien dem Grundkonzept des Internet, der ungesteuerten, individuellen Bereitstellung von Information, fast zu widersprechen, dass unter der Adresse <<http://www.archive.nrw.de>> mit einem Schlag allgemeine Informationen und Beständeübersichten zu 430 Archiven bereitgestellt wurden.⁷

Ungeachtet aller Diskussionen um das Konzept von *archive.nrw.de* erweisen die Nutzungszahlen das von der Deutschen Forschungsgemeinschaft

⁷ Über Motive und Hintergründe der seinerzeitigen Entwicklung ist genügend publiziert worden, so dass eine weitere Thematisierung an dieser Stelle unnötig ist. Vgl. zugleich auch stellvertretend für andere Veröffentlichungen Wilfried Reininghaus und Frank M. Bischoff: Archive in Nordrhein-Westfalen im Internet. Bericht über ein von der Deutschen Forschungsgemeinschaft unterstütztes Pilotprojekt. In: Der Archivar 51 (1998), Sp. 411–426.

geförderte Portal als Erfolg. Mit rund eine Million Hits pro Monat⁸ darf die Website als sehr gut besucht gelten. Wo liegen die Ursachen eines solchen Erfolgs?

Die wichtigste Voraussetzung ist sicherlich, dass hier Inhalte angeboten werden. Mit 430 Beständeübersichten dürfte Nordrhein-Westfalen noch heute das Bundesland sein, das seine Archivlandschaft in der Fläche am breitesten beschreibt. Auswertungen der nachgefragten Inhalte zeigen sehr deutlich, dass die Beständeübersichten im Zentrum des Interesses von Internetbesuchern stehen.⁹

Ein weiterer Grund besteht sicherlich in der Breite des Angebots. Die Archivlandschaft Nordrhein-Westfalen ist in dem Portal *archive.nrw.de* ausgesprochen dicht dokumentiert. In einem archivischen Internet-Verbundsystem kann der Nutzen von Informationsangeboten folglich maximiert und ein informationeller Mehrwert für den Nutzer wie auch für die Archive geschaffen werden. Einerseits wird es den beteiligten Archiven ermöglicht, sich mit eigener Homepage und eigener Beständeübersicht zu präsentieren. Andererseits kann der Benutzer eine archivübergreifende Recherche durchführen und dabei auch auf unerwartete Treffer stoßen. Im Unterschied zu den gängigen Suchmaschinen genügt die Qualität der Rechercheergebnisse in einem Fachinformationssystem einem gewissen Mindeststandard, so dass der Benutzer zugleich vor einer ‚Vergoogelung‘ geschützt ist.

Ein weiterer Vorteil eines Portals mag schließlich auch darin liegen, dass eine einheitliche Rechercheoberfläche für die Ermittlung von archivischen Informationen genutzt werden kann und dass die vorhandenen Informationen in einer fachlich angemessenen, einheitlichen Struktur organisiert sind.

Die Errichtung von Portalen ist inzwischen auch in einer Reihe anderer Bundesländer erfolgt. Portale erleichtern den Zugriff auf einzelne Archive oder Gruppen von Archiven erheblich. Insofern kann hier den Sachverständigen der Europäischen Kommission nur zugestimmt werden, wenn diese die Auffassung vertreten, dass Portale in Zukunft verstärkt eingerichtet oder ausgebaut werden sollten, vor allem im Interesse von Benutzern, aber

⁸ Vgl. die Nutzungsstatistik unter Adresse: <http://www.archive.nrw.de/statistik/default.html>, eingesehen am 26.5.2006.

⁹ Vgl. Frank M. Bischoff: Das Projekt *Archive in NRW* im Internet. Nutzung und Fortschreibung. In: *Archivpflege in Westfalen und Lippe* 53 (2000). S. 13–19, und ders.: Die Ausweitung archivischer Informationsvermittlung im nordrhein-westfälischen Internetportal und der Beitrag der Archive. In: *Archivpflege in Westfalen und Lippe* 57 (2002). S. 50–56.

auch im Interesse der Archive. Wenn der Rat der Europäischen Union dieses Gateway von den Nationalarchiven betreuen lassen möchte, dann wird damit den Größenverhältnissen im Archivwesen Rechnung getragen. Diesen Aufwand können nur große Archivverwaltungen auf sich nehmen; kleine Archive wären damit auf lange Sicht überfordert.

Vor dem Hintergrund der bestehenden Regionalportale in Deutschland und der Forderung nach einem EU-Archivportal wird eine Lücke offenkundig: Es fehlt ein Archivportal, das auf nationaler Ebene die Angebote der Archive in Deutschland bündelt. Zwar wird in dem Sachverständigenbericht der EU-Kommission ausgeführt, dass in Deutschland die Einrichtung eines nationalen Archivportals in Bearbeitung sei.¹⁰ Jedoch haben sich diese Arbeiten zunächst noch einmal verzögert. Ein deutsches Archivportal könnte die archivischen Internetressourcen sowohl gegenüber internationalen Archivverbänden als auch gegenüber weiter gefassten nationalen oder internationalen Informationsverbänden auf dem Gebiet des kulturellen Erbes gebündelt repräsentieren. Eine derartige Bündelung, die auf den Homepages und regionalen Portalen der deutschen Archive aufbauen müsste, würde auf nationaler Ebene eine Informationsmasse vermitteln, deren Bedeutung Benutzern und anderen Gateway-Betreibern unmittelbar augenfällig würde.

2. Schnittstellen, Austausch- und Präsentationsformate

Das Beispiel von *archive.nrw.de* lehrt auch, dass ein Portal ohne Informationszuwachs ab einem bestimmten Zeitpunkt zu stagnieren droht. Die Aufnahme neuer Archive, die Ausweitung oder Überarbeitung von Beständeübersichten sowie die Bereitstellung von weiteren, für Benutzer relevanten Informationen ist seit Einrichtung des Systems sehr langsam erfolgt. Seit Ende des Jahres 2002 lässt sich beobachten, dass sich die Nutzungsfrequenz auf die bereits erwähnten eine Million Hits pro Monat eingependelt hat, de facto also stagniert, wenngleich auf hohem Niveau. Ein Portal benötigt daher auch Informationszuwachs, um sich langfristig zu behaupten. In Nordrhein-Westfalen werden die Weichen seit einiger Zeit in diese Richtung ge-

¹⁰ Report, wie Anm. 5, S. 61.

stellt: Das bisher auf Beständeübersichten beschränkte Portal wird gerade um die Findbuchebene erweitert.

Ein derartiges Portal ist mit Hilfe von Fördermitteln der DFG vor wenigen Jahren in einem anderen Bundesland zum Abschluss gebracht worden. In Mecklenburg-Vorpommern wurde unter Federführung des Universitätsarchivs Greifswald das *Ariadne*-Portal geschaffen,¹¹ in dem Archive ihre Beständeübersichten und Findbücher im Internet bereitstellen können.

Die im *Ariadne*-Portal vorgesehenen Findmittelstrukturen entsprechen durchaus dem in Deutschland üblichen Archivstandard. Ähnlich wie *archive.nrw.de* kennt *Ariadne* Beständeübersichten mit in die jeweilige Tektonik eingebetteten Beständen sowie Findbücher mit Klassifikationen, an deren untersten Zweigen jeweils die Archivalieneinheiten hängen. *Ariadne* unterscheidet vier Archivalienarten, neben Sachakten auch Urkunden, Personalakten und Karten.

Die Beteiligung an dem *Ariadne*-Portal setzt in der Praxis voraus, dass die Archive auch die *Ariadne*-Erschließungssoftware einsetzen.¹² Diese kann als lokale Anwendung in einem lokalen Netz, aber auch als zentrale Anwendung auf dem *Ariadne*-Rechner des Rechenzentrums der Universität Greifswald, genutzt werden. Mit diesem Ansatz ist das *Ariadne*-Portal im Prinzip dem nordrhein-westfälischen Vorgehen gefolgt. Auch in Nordrhein-Westfalen haben die Archive eine lokale Software zur Pflege ihrer Beständeübersichten erhalten. Hier ist also die Frage berührt, ob eine Grundvoraussetzung zur Errichtung eines Portals darin bestehen muss, dass alle beteiligten Archive dieselbe Software nutzen.

Blicken wir nochmals für einen Moment in die Entwicklungsphase von *archive.nrw.de* im Jahr 1997 zurück. Mit dem Vorhaben, ein zentrales Archivportal einzurichten, dessen Inhalte dezentral gepflegt werden sollten, stellte sich zugleich die Frage, wie die Beständeinformationen verschiedener Archive in einer inhaltlich und technisch einheitlichen Struktur auf einen zentralen Internetserver geladen werden können. Es drängte sich der Gedanke auf, eine Bestände-Software zu entwickeln und alle beteiligten Archive zu verpflichten, dieses Werkzeug für die Pflege ihrer Beständeüber-

¹¹ Archive Information & Administration Network, Adresse: <http://ariadne.uni-greifswald.de> (letzte Einsichtnahme am 26.05.2006).

¹² An dieser Stelle sei Herrn Dr. Matthias Manke, Landeshauptarchiv Schwerin, für den freundlichen Hinweis auf die bestehende ASCII-CSV-Importschnittstelle von *Ariadne* gedankt, die es ermöglicht, ohne den Einsatz der Software Findmittel-Daten in *Ariadne* zu importieren.

sichten zu nutzen. Die Software wurde den beteiligten Archiven kostenlos zur Verfügung gestellt. Die Beständeübersicht des jeweiligen Archivs wurde bei der Übergabe der Software als Datenbank bereitgestellt.

Bei der Beurteilung dieses Vorgehens ist zu berücksichtigen, dass vor neun Jahren die Ausgangssituation eine andere war als heute. Es existierten keine marktgängigen Systeme zur Pflege von Beständeübersichten. Das heißt, dass mit der Bereitstellung eines solchen Tools weder eine Umgewöhnung der Archive noch eine Marktverzerrung verbunden war. Für die beteiligten Archive gab es nur Vorteile, für Softwarehersteller mangels kommerzieller Produkte keine Nachteile.

Das ist heute anders. Erstens gibt es inzwischen Hersteller, die auch Systeme zur Pflege von Beständeübersichten in ihrer Softwarepalette anbieten. Vor allem aber zielen die heutigen Bestrebungen auf die Findbuchebene, ein Bereich, für den in Deutschland seit Jahren schon eine Reihe von Produkten am Markt erhältlich ist. Und diese Produkte werden ebenfalls bereits seit Jahren in Archiven zur Verzeichnung und Erstellung von Findbüchern eingesetzt. Insofern kann die Verpflichtung zur Nutzung eines bestimmten Systems zur Bestückung eines Internet-Portals weder für die Archive noch für den Softwaremarkt – letztlich profitieren die Archive von einer Marktvielfalt – ein probates Verfahren sein.

Damit wird ein gravierendes Defizit innerhalb des deutschen Archivwesens offenbar. Im Gegensatz zu den Bibliotheken haben sich die deutschen Archive in der Vergangenheit nicht in ausreichendem Maße mit der elektronischen Kommunizierbarkeit und Austauschbarkeit ihrer Erschließungsinformationen beschäftigt. Solange Findmittel nur in Buchform oder als Karteikarten existierten, bestand angesichts der Einzigartigkeit von Archivgut dazu vielleicht auch wenig Anlass. Gewisse Quasi-Standards wurden über die Ausbildung und über archivfachliche Publikationen verbreitet. Mit der Einrichtung von Informationsverbänden fällt dieses Defizit aber auf die Archive zurück, da elektronische Systeme nicht unscharf umrissene Quasi-Standards, sondern nur eindeutig definierte Strukturen und Elemente verarbeiten können. Es hat heute den Anschein, als seien Lösungen in einem eher durch Individualität geprägten Metier schwer zu finden.

Als das Hauptstaatsarchiv Düsseldorf vor fünf Jahren das Projekt zur Retrokonversion von Findmitteln startete, stellten sich ähnliche Fragen. Das Projekt zielte in erster Linie auf die Entwicklung eines Werkzeugs, das die automatische oder halbautomatische Konversion maschinengeschrie-

bener oder gedruckter Findmittel in eine strukturierte digitale Umgebung leisten sollte. Schnell war klar, dass dies nur die eine Hälfte der Aufgabe sein konnte. Die andere musste darin bestehen, die in ein digitales Format konvertierten Findbücher mit relativ geringem Aufwand in verschiedene Zielsysteme zu portieren. Die Lösung schien in der Schaffung eines XML-gestützten Austauschformates zu liegen, das in seinen Strukturen und Elementen eindeutig definiert war.¹³ Wenn in bestehenden Erschließungsdatenbanken dieses Austauschformat als Importformat definiert und eingebunden würde, könnten die Archive die konvertierten Findbücher mühelos in ihre Software laden.

Das im Rahmen des Düsseldorfer Retrokonversionsprojekts entwickelte Austauschformat ist folglich darauf ausgerichtet, jede isolierbare oder in Findbüchern inhaltlich oder optisch als eigenständig erkennbare Information als separates Element in einem XML-Dokument abzulegen, um auf diese Weise jeglichen Informationsverlust und etwaige Nach- und Mehrarbeit zu vermeiden. Wie der Name bereits aussagt, dient es dem verlustfreien Austausch von Erschließungsinformationen zwischen verschiedenen Systemen. Es ist nicht zur Internet-Präsentation von Erschließungsinformationen entwickelt worden.

Es entspricht durchaus den auch das deutsche Archivwesen prägenden föderalen Strukturen, wenn inzwischen weitere Einrichtungen XML-DTDs für Findmittel entwickelt haben. Dies ist nicht nur beim *Ariadne*-Portal der Fall, das auch den Import von XML-Dokumenten vorsieht und zu diesem Zweck eine eigene DTD definiert hat. Auch der von der Archivschule Marburg in Kooperation mit dem Bundesarchiv entwickelte Findbucheditor MidosoXML¹⁴ verfügt über eine eigene DTD. MidosoXML ist zugleich das einzige in Deutschland existierende Verzeichnungsprogramm, das derzeit mit einer funktionierenden EAD-Schnittstelle ausgestattet ist. Mit MidosoXML kann sozusagen per Knopfdruck ein EAD-konformes XML-Findbuch exportiert werden. MidosoXML liegt unter anderem die Philosophie frei austauschbarer Findmittelinformationen zugrunde, was in dem System über verschiedene Import- und Exportroutinen realisiert wird. Es wäre

¹³ Vgl. Stefan Przigoda: Das Ziel- und Austauschformat als universelle Findbuch-DTD, Adresse: <http://www.archive.nrw.de/dok/tagung-retro/11-Przigoda.pdf> (letzte Einsichtnahme am 26.05.2006).

¹⁴ Vgl. Adresse: <http://www.archivschule.de/content/26.html> (letzte Einsichtnahme am 26.05.2006).

wünschenswert, wenn dieser Grundgedanke allmählich auch in andere archivistische Erschließungssysteme Eingang finden würde.

Die ISO-Norm ‚Encoded Archival Description‘ (EAD) findet bisher in Deutschland außerhalb des Bundesarchivs wenig Beachtung. Aus der Perspektive eines verlustfrei arbeitenden Austauschformats mag das seine Berechtigung haben, da EAD manche Besonderheiten von historischen Archivalientypen nicht berücksichtigt. Zudem umfasst EAD Elemente, die die deutsche Erschließungstradition kaum kennt. Formuliert man die Ansprüche aber anders und fragt, ob EAD als Präsentationsformat für eine Internet-Präsentation von Findmitteln geeignet sein könnte, muss die Antwort positiv ausfallen. Informationen, die als Kernelemente einer Erschließung gelten dürfen, sind in EAD berücksichtigt. Insbesondere verfügt EAD über eine hierarchische Struktur, die auch der deutschen Erschließungstradition von Tektonikebenen und Beständen, Findbüchern mit Klassifikationen und Archivalieneinheiten entspricht.

Was also wird benötigt, um Erschließungsinformationen im Internet in geeigneter Weise darzustellen? Das nordrhein-westfälische Beständeportal differenziert neun verschiedene Elemente. Auf dem Internetserver werden die in diesen Elementen erfassten Informationen aber zusammengefügt und in einem einzigen Feld abgelegt. Mag man das auch heute als zu grobstrig empfinden, so ist die Zahl der sinnvoller Weise zu differenzierenden Informationen, damit der Benutzer gezielt bestimmte Informationen recherchieren und abfragen kann, sicherlich eng begrenzt und müsste unterhalb von zehn Elementen liegen. Mehr würde eine Überforderung der meisten Benutzer darstellen, würde die Recherche erheblich verkomplizieren und wäre damit kontraproduktiv. Deutsche Archivare täten sicher gut daran, wenn sie von der ihnen offensichtlichen Vielfalt von Erschließungsdetails im Interesse von Benutzern abstrahieren und sich bei der Internetpräsentation von Findmitteln auf ein überschaubares Raster beschränken würden. Weniger kann hier mehr sein.

Ungeachtet der Skepsis deutscher Fachkolleginnen und -kollegen gegenüber EAD sei an dieser Stelle nachdrücklich hervorgehoben, dass EAD das leisten kann und als Präsentationsformat im Internet sehr gut geeignet ist. Einmal mehr sei auf den Sachverständigenbericht über die Archive in der erweiterten Europäischen Union verwiesen, der EAD als internationalen

Archivstandard empfiehlt und eine Reihe von europäischen Internetressourcen zitiert, die auf EAD aufbauen.¹⁵

Bevor man also eine Findmittel-Kodierung nach EAD ablehnt, sollte gründlich geprüft werden, welche erheblichen Mängel gegen seinen Einsatz als Präsentationsformat sprechen. Da die Deutsche Forschungsgemeinschaft in den vergangenen Jahren im Rahmen zweier Projekte bereits die Entwicklung zweier unterschiedlicher Findmittel-DTDs finanziert hat, sollte auch kritisch geprüft werden, ob die Notwendigkeit zur Entwicklung einer dritten wirklich besteht. Auch hier kann weniger mehr sein.

Es sei nochmals unterstrichen, dass eine erfolgreiche Einrichtung von zentral zugänglichen, aber inhaltlich dezentral gepflegten Archivportalen in Zukunft die Existenz von anerkannten und eindeutigen Schnittstellen erfordert. Die dazu notwendigen Sachentscheidungen müssen von Archiven getroffen und am besten von entsprechenden Empfehlungen archivischer Gremien begleitet werden. Ohne derartige Schnittstellen werden es Archivportale schwer haben, eine kritische Masse von Findmitteln aus einer möglichst großen Zahl von Einrichtungen zusammenzuführen.

3. Internetfindmittel und Retrokonversion

Archivportale bieten Archiven wie Benutzern einen hohen Komfort für die Bereitstellung von Erschließungsinformationen im Internet, Schnittstellen sind eine wichtige Voraussetzung für die Kommunikation von Erschließungsdaten zu Portalen. Bleibt der Frage nachzugehen, wie in möglichst kurzer Zeit eine möglichst große Zahl von Findmitteln digital aufbereitet werden kann.

Archivische Findmittel sind und bleiben Generationenwerk und können in absehbarer Zeit nicht neu erstellt werden. Vor diesem Hintergrund müssen die bereits in den Bibliotheken eingesetzten Strategien der Retrokonversion auch in den Archiven zum Einsatz gelangen. In Anlehnung an die bibliothekarische Terminologie wird im Folgenden unter ‚Retrokonversion‘ die Umsetzung älterer Findmittel in ein digitales Format verstanden. Die Übertragung erfolgt dabei im Wesentlichen unverändert, ist also nicht von nennenswerten Nach- oder Neuverzeichnungsarbeiten begleitet.

¹⁵ Report, wie Anm. 5, S. 61 f.

In deutschen Archiven wurden in den vergangenen Jahren unterschiedliche Ansätze zur Retrokonversion verfolgt. Das Hauptstaatsarchiv Düsseldorf hat ein Werkzeug entwickelt, das es ermöglicht, aufgrund von inhaltlichen und Layout-Merkmalen die in maschinengeschriebenen oder gedruckten Findmitteln enthaltenen Informationen zu identifizieren, zu zerlegen und strukturiert in einem XML-Dokument abzulegen. Das Werkzeug ist mit Fördermitteln der DFG entwickelt worden und kann folglich von deutschen Archiven nachgenutzt werden.¹⁶

Die Stiftung ‚Archiv der Parteien und Massenorganisation der DDR‘ im Bundesarchiv hat in den letzten Jahren mehrfach Findmittelretrokonversionen im Rahmen von Fremdaufträgen abgewickelt. Die Aufträge wurden zum Teil in Rumänien durch händische Erfassung in einem strukturierten Format abgewickelt. Sofern es sich um handschriftliche Findmittel handelt, ist die händische Erfassung zugleich die einzige Möglichkeit der Überführung in ein digitales Format.

Die hessischen Staatsarchive schließlich haben mit eigenem Personal unterstützt durch Projekt-Hilfskräfte ihre maschinengeschriebenen Findmittel in den letzten Jahren gescannt, mit OCR bearbeitet, die elektronischen Texte dann händisch strukturiert und in ihre Datenbank *HADIS* eingespeist. In Wiesbaden und in Darmstadt ist der größte Teil der Arbeiten zum Abschluss gelangt.

Die drei genannten Beispiele zeigen, auf welchen Wegen klassische Findbücher schnell und kostengünstig in elektronische Formate umgewandelt werden können. Welche Methode die geeignete ist, hängt von der Qualität und vom Umfang des Ausgangsmaterials ab. Das muss sicherlich im Einzelnen geprüft werden. An dieser Stelle sei lediglich unterstrichen, dass eine Retrokonversion die einzige Möglichkeit ist, Findmittel in großer Zahl, kostengünstig und schnell im Internet bereitzustellen. Es wäre fatal, wenn der Qualitätsanspruch dabei zu hoch geschraubt würde. Hier ist vielmehr eine pragmatische Herangehensweise gefordert. Ein Findbuch der 60er Jahre, das lediglich eine Kurztitelverzeichnung bietet, kann im Internet für viele Benutzer von großem Wert sein. Wird es zurückgehalten, weil zuvor eine Neuerschließung erfolgen soll, dürfte der Bestand angesichts durchaus beträchtlicher Erschließungsrückstände in vielen Archiven auf Jahre hinaus

¹⁶ Vgl. den Abschlussbericht zum DFG-Projekt *Entwicklung von Werkzeugen zur Retrokonversion archivischer Findmittel* unter Adresse: http://www.archive.nrw.de/dok/dfg_abschluss/abschlussbericht_retrokonversionsprojekt.pdf (letzte Einsichtnahme am 26.05.2006).

für den Internetbenutzer verloren sein. Archive, die den notwendigen Pragmatismus zu derartigen Veröffentlichungsentscheidungen aufbringen, sollten darin gestützt und gefördert werden.

Inzwischen gibt es seitens der Deutschen Forschungsgemeinschaft eindeutige Aussagen, die deutschen Archive in ihren Bemühungen um die Retrokonversion von Findmitteln in den nächsten Jahren zu unterstützen. Das ist ermutigend und lässt hoffen, dass die deutschen Archive in fünf bis zehn Jahren einen großen Schritt auf dem Weg der Online-Bereitstellung von Erschließungsinformationen getan haben werden. Das setzt ein rasches, zielorientiertes Handeln der Archive voraus. Es muss bei der Retrokonversion darum gehen, schnell und kostengünstig eine große Masse von Findmitteln zu bearbeiten und im Internet bereitzustellen. Dass es sich dabei um frei zugängliche Bestände bzw. Findmittelinformationen handeln muss, ist selbstverständlich. Hier aber die Frage der Bedeutung eines Bestandes für den einen oder anderen Aspekt der geschichtlichen Forschung einzumischen, würde die Gefahr unerträglicher Verzögerungen in sich bergen. Insofern sollte die Retrokonversion von Findmitteln zwar planmäßig, aber auch pragmatisch betrieben werden.

4. Zusammenfassung

Zum Abschluss seien die wesentlichen Aussagen nochmals in kurzen Thesen zusammengefasst:

- Portale machen die Archivlandschaft einer Region oder einer Sparte transparent und bieten sowohl Benutzern wie Archiven einen informationellen Mehrwert. Ihre Einrichtung oder Erweiterung sollte daher vorangetrieben werden. Es besteht dringender Bedarf, ein nationales Gateway einzurichten, das die deutsche Archivlandschaft gegenüber nationalen oder internationalen Informationsverbänden auf dem Gebiet der historischen Forschung und des kulturellen Erbes bündelt und vermittelt.
- Zur Kommunikation von Erschließungsinformationen werden Schnittstellen benötigt. Die deutschen Archive müssen hier längst überfällige Entscheidungen treffen und sich rasch auf Standards einigen. Es muss sich nicht zwangsläufig um eine einzige Austausch- oder Präsentationschnittstelle handeln, notwendigerweise aber um wenige, wobei im

Hinblick auf einen internationalen Austausch EAD inzwischen unabdingbar dazugehört. Ohne definierte Schnittstellen werden Findmittel auch in Zukunft nur mit hohem Kostenaufwand kommunizierbar sein oder im Zweifelsfall isoliert in proprietären Umgebungen verbleiben.

- Die Retrokonversion archivischer Findmittel sollte zügig und nach sehr pragmatischen Gesichtspunkten vorangetrieben werden, um Internetbenutzern möglichst rasch eine kritische Masse von Findmitteln zur Verfügung stellen zu können. Eine Förderung durch die Deutsche Forschungsgemeinschaft ist inzwischen zugesagt und muss jetzt von den deutschen Archiven in Ergebnisse umgesetzt werden.

Geschichtswissenschaft auf dem Weg zur E-History?

Angelika Schaser

Fünf Themenkreise standen auf der Tagung *Forschung in der digitalen Welt. Sicherung, Erschließung und Aufbereitung von Wissensbeständen im Mittelpunkt*. Die ‚Probleme und die Chancen der Digitalisierung‘ (Thaller, Sahle, Heller/Vogeler), die ‚Macht der Bilder und die Beschleunigung der wissenschaftlichen Diskussion‘ über die Visualisierung archäologischer Befunde (Schäfer), ‚Quelleneditionen‘ im Bereich der mittelalterlichen und frühneuzeitlichen Geschichte (Sarnowsky, Rübsamen/Kuczera, Schildt, Staecker), eine ‚Online-Enzyklopädie‘ (Lorenz/Krüger) und ‚Archive in der digitalen Welt‘ (Hering, Heckmann, Bischoff/Schäfer) wurden präsentiert und diskutiert. Nach der Vorstellung der Projekte und wichtiger theoretischer, methodischer und praktischer Überlegungen für die Aufbereitung, Erhaltung und Bereitstellung von historischem Material und Publikationen für die Forschung bleibt danach zu fragen, welche Schlussfolgerungen aus dem Präsentierten gezogen werden können.

Aus der Fülle der Themenbereiche, die dabei berührt wurden, sollen hier vier Punkte herausgestellt werden. Vorauszuschicken bleibt, dass die stark technik-orientierten Beiträge technisch weniger versierten Zuhörern deutlich machten, dass auf diesem Gebiet inzwischen ein Expertenwissen existiert, das nicht einfach zu erschließen ist. Ähnlich wie für die gebräuchlichen Abkürzungen des 16. bis 18. Jahrhunderts wünschte man sich hier als Nichtexpertin für solche Vorträge ein Abkürzungsverzeichnis und einige Erläuterungen. Gleichzeitig wurde jedoch auch dem naiven Benutzer digitaler Angebote eindringlich vor Augen geführt, dass eine Pioniergruppe hoch motivierter und nicht selten durch Doppelqualifikationen ausgewiesener Historiker Quellen digitalisiert, Datenbanken und Fachportale

aufbaut sowie neue Medien für den Unterricht herstellt, während andere dieser Entwicklung noch distanziert bis ablehnend gegenüberstehen. Dass die Digitalisierung allerdings längst die Forschung beeinflusst, ohne dass ein eindeutiges Votum von Historikern und Archivaren zu dieser Entwicklung bislang vorläge, haben auch die 2004 in Hamburg abgehaltene Tagung zum Thema *Im Netz des Positivismus? Vom Nutzen und Nachteil des Internet für die historische Erkenntnis* sowie ein im selben Jahr durchgeführter Workshop in Fribourg (Schweiz) deutlich gemacht.¹

1. Zum Wandel der Geschichtswissenschaft in der digitalen Welt

In seinem einführenden Eingangsvortrag hat Thaller diesen schleichenden und bislang wenig reflektierten Veränderungsprozess, der durch das WWW und die neuen Medien ausgelöst wurde, angesprochen. Diese Veränderung bietet seiner Meinung nach zum ersten Mal die Möglichkeit zu einem radikalen Wandel der Geschichtswissenschaft. Ein solcher ‚Paradigmenwechsel‘, der diesen Namen wie keine andere Richtungsänderung innerhalb der historischen Wissenschaft verdienen würde, könnte eine Dekanonisierung und die Auflösung asynchroner Arbeitsvorgänge mit sich bringen. Dies würde jedoch wiederum den radikalen Wandel des Selbstverständnisses von Archivaren, Bibliothekaren und Forschern sowie einen kooperativen Arbeitsstil erfordern. Die vorgestellten Großprojekte zur Digitalisierung mittelalterlicher und frühneuzeitlicher Quellen zeigen jedoch einen gegenteiligen Trend, die Kanonisierung wird mit diesen Unternehmungen eher verstärkt als abgebaut. Während auf der einen Seite Geschichtsschreibung also auch unter neuen technischen Möglichkeiten weitgehend alten Mustern folgt, zeigt Thaller auf, dass auch das Festhalten an ehrwürdigen Traditionen letztlich nicht die bereits in Gang gesetzte Änderung der Welt der Wissenschaften aufhalten können wird. So werden zwar die benutzten Quellen und die eingearbeitete Literatur auch heute noch in einem so genannten wissenschaftlichen Apparat in Form von Fußnoten be-

¹ Vom Nutzen und Nachteil des Internet für die Historische Erkenntnis. Version 1.0. Hg. von Angelika Epple und Peter Haber (Geschichte und Informatik, Histoire et Informatique 15). Zürich 2005.

legt.² Diese Methode war ursprünglich auf einen kleinen elitären Gelehrtenzirkel zugeschnitten, der mit den im 19. Jahrhundert edierten Quellen und der Forschungsliteratur vertraut war. Angesichts der Massenproduktion von wissenschaftlichen historischen Texten können die in den Fußnoten angegebenen Belege von uns nicht mehr vollständig nachvollzogen, geschweige denn kontrolliert werden. Kritische Studierende stellen deshalb nicht zu Unrecht Lehrenden häufig die Frage nach der Funktion und der Bedeutung von Fußnoten. Hier zeigt die von Staecker aufgestellte These, dass gerade in der Vernetzung die Chance der Digitalisate liegt, eine interessante Perspektive auf. Eine Verlinkung in der Fußnote auf die entsprechende Textstelle in den Quellen bzw. in der Sekundärliteratur würde die Argumentation sowie Quellenbasis wissenschaftlicher Arbeiten in ganz neuer Weise nachvollziehbar machen und nachdrücklich zum wissenschaftlichen Dialog auffordern.

Wie sehr das WWW unsere Wissenschaft verändert hat, zeigt der Beitrag von Lorenz/Krüger, der leider nicht für den Druck zur Verfügung stand. Die Einträge in der Online-Enzyklopädie werden von Wissenschaftlern wie von Studierenden gleichermaßen häufig genutzt, obwohl die Frage nach der Autorenschaft und der Qualitätssicherung der dort abrufbaren Texte ungeklärt ist. Bei diesem Projekt, das mit aufklärerischer und demokratischer Zielsetzung angetreten ist, taucht wie bei den anderen vorgestellten Großprojekten die Frage auf, wie und von wem wissenschaftliche Standards definiert werden sollen und ob ohne Institutionalisierung eine funktionierende Qualitätssicherung gewährleistet werden kann.

2. Nachhaltigkeit/Finanzierbarkeit der Projekte

Nahezu in allen Beiträgen kam die Frage nach der nachhaltigen Sicherung der digital bereit gestellten Materialien und der Zukunft der vorgestellten Projekte auf. Was will sich die Wissenschaftspolitik in Zukunft leisten, was kann sie sich leisten? Wieweit ist die Kommerzialisierung nötig? Bringt sie nur Nachteile mit sich oder birgt sie auch Chancen? Kann bei der anwachsenden Quantität von Digitalisierung und Katalogisierung eine gleichbleibende Qualität garantiert werden? Wenn ‚work in progress‘ sinnvoller Wei-

² Anthony Grafton: Die tragischen Ursprünge der deutschen Fußnote. München 1998, bes. S. 13–47.

se veröffentlicht werden soll, wann ist dann eine elektronische Edition abgeschlossen? Wer erklärt sie dazu? Wenn viele mitarbeiten können und sollen, wer behält den Überblick? Wer schaltet Texte wann frei? Wie können Angebote zusammengeführt, Doppelarbeiten vermieden werden? Wie von Welck angekündigt, betrachtet die (Hamburger) Bildungspolitik die Bereitstellung und die Bewahrung von Wissen als eine zentrale politische Aufgabe. Kooperationen werden für die Bewältigung dieser Aufgabe für notwendig erachtet, über das Finanzvolumen für derartige Vorhaben wurde nichts verraten.

Sahle und Bischoff führen in diesem Themenbereich einige Grundüberlegungen zu Techniken und Standards für die Erschließung historischer Dokumente vor. Bischoff gibt die Devise aus, man müsse planmäßig, aber pragmatisch vorgehen. Die Archive scheinen bezüglich der Vernetzung – trotz der Föderalismusproblematik – in dieser Hinsicht einen Schritt weiter zu sein als die universitäre Geschichtsschreibung. Sahle weist darauf hin, wie dringend notwendig für die gewünschte Vernetzung eine klare Modularisierung des Informationsraums ist, die Input, Datenhaltung und Output klar voneinander trennt. Solche Vernetzungen werden im regionalen, im deutschen oder im internationalen Rahmen trotz aller Hindernisse von den vorgestellten mittelalterlichen und dem frühneuzeitlichen Projekt bereits hergestellt.

Mit den von Thaller genannten circa 90 Pilotprojekten in Deutschland, die eine systematische Umsetzung großer Bestände von archivalischen Quellen und Bibliotheksbeständen anstreben, existieren also bereits vielversprechende, inspirierende Ansätze für die weitere Vernetzung. Ob es dabei zur Revolutionierung der Geschichtswissenschaft kommen wird, muss sich noch zeigen. Wer wird diese Datenbanken und Digitalisate nutzen, wer wird mit den Anbietern dieser Wissensräume in den interaktiven Dialog treten? Wird es zu einem neuen, kooperativen Arbeitsstil kommen? Wird neben den ‚klassischen‘ Quellen auch neues Material der Forschung online zugänglich gemacht, das von der Geschichtswissenschaft auf diesem Weg leichter entdeckt werden kann? Werden die aufbereiteten Daten und Texte auch für breitere Kreise in Zukunft lesbar und benutzbar sein? Auf welchen Rechnern werden die frühen Digitalisate zu lesen sein? Am privaten PC oder nur mehr im Museum für Technik?

3. Einzug eines neuen Positivismus?

Eine in diesem Zusammenhang schon oft gestellte Frage wurde auch auf dieser Tagung diskutiert. Die von Heller und Vogeler mit großem Nachdruck vertretene Forderung, Informatikverfahren für die historische Anwendung im größerem Maßstab nutzbar zu machen, wirkt zunächst nicht zuletzt durch die von den Referenten deutlich gemachte Begeisterung für ihre vorgestellte Suche nach Problemlösungen faszinierend. Sicher bieten digitalisierte Texte der Forschung leichtere Zugänglichkeit und phantastische Möglichkeiten für erzähltheoretische und philologische Fragestellungen. Wieweit ist jedoch die Informationsrecherche automatisierbar? Ist Text nicht doch mehr als eine Zeichenkette? Wird mit der angestrebten Form der Textaufbereitung nicht einem neuen Positivismus der Weg bereitet, wenn suggeriert wird, dass Historiker mit speziell für die Geschichtswissenschaft entwickelten Suchmaschinen in den aufbereiteten Texten ‚sachlichen Inhalt‘ finden könnten? Da Informationen niemals verlustfrei von einem Medium in ein anderes transportiert werden können, ist auch danach zu fragen, welche Informationen bei der Digitalisierung verloren gehen werden. Wieweit kann, wie Hering fragt, die notwendige Kontextualisierung der Dokumente geleistet werden? Wird in absehbarer Zeit für die Forschung nur noch das existieren, was im WWW zu finden ist?

Ähnliche Fragen tauchten auch nach dem Beitrag von C. Schäfer auf: Wie verträgt sich die Macht der Bilder – die im Übrigen das Geschichtsbewusstsein breiter Bevölkerungsschichten schon heute weit eindringlicher prägt als gelehrte Texte – mit einer differenzierten Darstellung von Geschichte? Wie kann man eindrucksvolle Visualisierungen mit Problematisierung und Differenzierung verbinden? Auch hier wird zu beobachten sein, in welche Richtung die neuen Recherchemöglichkeiten und -strategien sowie Präsentationsformen die Geschichtswissenschaft verändern werden.

4. Archive als „Gedächtnis der Gesellschaft“

Nicht nur die Geschichtswissenschaft, auch die Archive stehen vor technischen Herausforderungen und mentalen Umstellungen bei der Aufnahme von Archivgut. Die Entwicklung der Verwaltungen hin zum E-Government verlangen nach neuen Wegen der Archivierung. Die Arbeitsprozesse wer-

den sich verändern, Fragen der Benutzbarkeit und des Urheberrechts müssen bezüglich der neuen technischen Möglichkeiten berücksichtigt werden. Nicht zuletzt ist eine Anpassung und Weiterentwicklung der historischen Hilfswissenschaften an diesen Prozess erforderlich. Ein erstes ehrgeiziges europaweites Ziel ist die Online-Bereitstellung aller Findbücher. Ein gemeinsames Portal für Archive, Bibliotheken und Museen in der Bundesrepublik ist in Vorbereitung. Ein Ziel bei der Digitalisierung wird die Schonung der Original-Bestände und die Steigerung der Effizienz bei der Erschließung und der Bereitstellung des Archivguts sein. Auch hier stellt sich die Frage nach einheitlicher Software und einheitlichen Standards bei der Aufbereitung der Archivalien, um den Informationsaustausch und Konvertierungen zu erleichtern und die Bestände nachhaltig zu sichern.

Bei der Benutzerfreundlichkeit wurden große Unterschiede zwischen den Archiven deutlich. Während das Geheime Staatsarchiv in Berlin seinem Namen alle Ehre macht, sind andere Archive eher der Open-Access-Bewegung verpflichtet. Archivare, so die Prognose Bischoffs, werden sich zu Informationsbrokern entwickeln, nicht zuletzt deshalb, weil Archivaren keine Zeit mehr für die inhaltliche Erschließung der Archivalien bleiben wird. Ob sich Archive, wie Thaller meint, von Reisezielen zu Anbietern von Ressourcen wandeln werden, wird nicht zuletzt von Entscheidungen zum Urheberrecht und zu den Benutzergebühren abhängen. Deutlich wurde in der Abschlussdiskussion, dass für die anstehenden Aufgaben eine enge Zusammenarbeit zwischen Archivaren und Historikern nötig ist, um die Archive für die Zukunft zu Orten des ‚gesellschaftlichen Gedächtnisses‘ zu entwickeln. Die Tagung lieferte für diesen wissenschaftlichen Austausch erste wichtige Anstöße.

Beitragende

BISCHOFF, Dr. Frank M., geb. 1959. Der Historiker und Archivar hat nach seinem Archivreferendariat in Detmold und Marburg seit 1996 als Dezerent, seit 1999 als Abteilungsleiter im Staatsarchiv Münster ältere Bestände, moderne Behörden und die Informationstechnologie betreut. Seit 2003 ist er Leiter der Archivschule Marburg, wo er unter anderem Archivwissenschaft, Verwaltungswissenschaft und Historische Hilfswissenschaften lehrt.

HECKMANN, Dr. Dieter, geb. 1955, ist ausgebildeter Gymnasiallehrer und Archivar. Als Leiter eines Referates am Geheimen Staatsarchiv Preußischer Kulturbesitz war er lange Jahre zuständig für die Sicherungsverfilmung. Er ist unter anderem durch eine Reihe von Einzelschriften ausgewiesener Fachmann für Historische Grundlagenforschung.

HELLER, Markus, M.A., geb. 1974. Nach seinem Studium der Geschichte, Anglistik und Wirtschaftsinformatik an der Universität Regensburg und mehrjähriger Tätigkeit als Projektleiter in der Informatikbranche derzeit Doktorand am Centrum für Informations- und Sprachverarbeitung (CIS) der Universität München, zudem Stipendiat von Siemens Com.

HERING, Dr. Rainer, geb. 1961. Der Archivar ist seit 1987 im Staatsarchiv Hamburg tätig und Leiter der Fachabteilung sowie Stellvertreter des Amtsleiters. Ab Oktober 2006 wird er die Leitung des Landesarchivs Schleswig-Holstein übernehmen. Er ist Vorsitzender des *Archive Committee* der *German Studies Association* (USA) und lehrt als Privatdozent für Neuere Geschichte am Historischen Seminar der Universität Hamburg.

KUCZERA, Dr. Andreas, geb. 1972, ist wissenschaftlicher Mitarbeiter der Deutschen Kommission für die Bearbeitung der *Regesta Imperii* e. V. bei der Akademie der Wissenschaften und der Literatur Mainz und dort unter anderem mit der fachinformatischen Weiterentwicklung des Online-Angebots der *Regesta Imperii* betraut.

- RÜBSAMEN, Dr. Dieter, geb. 1953, ist wissenschaftlicher Mitarbeiter der Deutschen Kommission für die Bearbeitung der *Regesta Imperii e. V.* bei der Akademie der Wissenschaften und der Literatur Mainz und dort unter anderem mit Aufbau und Betreuung des Online-Angebots betraut.
- SAHLE, Patrick, M.A., geb. 1968, ist wissenschaftlicher Mitarbeiter an der niedersächsischen Staats- und Universitätsbibliothek Göttingen. Er arbeitet dort in den Projekten *zvdd (Zentrales Verzeichnis digitalisierter Drucke)* und *OPAL (Online-Portal für digitalisierte Kulturgüter Niedersachsens)*.
- SARNOWSKY, Dr. Jürgen, geb. 1955, ist seit 1996 Professor für mittelalterliche Geschichte an der Universität Hamburg. Seine Forschungsschwerpunkte sind die Bildungs- und Wissenschaftsgeschichte, Ordens- und Hanse-Geschichte, preußische und englische Geschichte.
- SCHÄFER, Dr. Christoph, geb. 1961. Der Historiker ist seit 2003 Professor für Alte Geschichte an der Universität Hamburg. Neben seinen Forschungsschwerpunkten im Hellenismus, zur Sozial- und Wirtschaftsgeschichte des Imperium Romanum und zu antiker Schifffahrt interessiert er sich besonders für digitalen Wissenstransfer.
- SCHÄFER, Dr. Udo, geb. 1959. Der Jurist und Archivar ist nach seiner Tätigkeit bei der staatlichen Archivverwaltung Baden-Württemberg seit 2001 Amtsleiter des Staatsarchivs der Freien und Hansestadt Hamburg. Daneben ist er Vorsitzender der Arbeitsgruppe *Archive und Recht* der Archivreferentenkonferenz des Bundes und der Länder.
- SCHASER, Dr. Angelika, geb. 1956, ist seit 2001 Professorin für Neuere Geschichte an der Universität Hamburg und leitete von 2003 bis 2005 das Teilprojekt *Medienkompetenz im Grundstudium* innerhalb des Projekts *Geisteswissenschaften im Internet*. Sie ist Leiterin der ‚Arbeitsstelle für die digitale Edition neuzeitlicher Selbstzeugnisse‘ am Zentrum ‚Geisteswissenschaften in der digitalen Welt‘ an der Universität Hamburg.
- SCHILDT, Dr. Bernd geb. 1948. Der Jurist ist Professor für Rechtsgeschichte und Bürgerliches Recht an der Ruhr-Universität Bochum. Schwerpunkt seiner wissenschaftlichen Arbeit ist die Erforschung der Rechtsprechung des Reichskammergerichts.
- STÄCKER, Dr. Thomas, geb. 1963, ist seit 1998 wissenschaftlicher Bibliothekar an der Herzog August Bibliothek Wolfenbüttel und Leiter der Abteilung Alte Drucke und Digitalisierung. Seine Arbeitsschwerpunkte sind ‚Digitalisierung‘ und ‚Altes Buch‘.

THALLER, Dr. Manfred, geb. 1950. Der Historiker ist seit 2000 Professor für Informatik in den Geisteswissenschaften an der Universität zu Köln. Zuvor war er am Max-Planck-Institut für Geschichte für das Arbeitsgebiet ‚Historische Fachinformatik‘ verantwortlich.

VOGELER, Dr. Georg, geb. 1969, ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Geschichtliche Hilfswissenschaften der Ludwig-Maximilians-Universität München. Nach seiner Dissertation aus dem Bereich der Amtsbücherkunde arbeitet er zur Digitalisierung von mittelalterlichen Urkunden und stieß 2004 die *Charters Encoding Initiative* an.

WELCK, Dr. Karin von, geb. 1947. Die Ethnologin und Honorarprofessorin der Universität Mannheim war Generalsekretärin der Kulturstiftung der Länder und ist seit März 2004 Präses der Kulturbehörde der Freien und Hansestadt Hamburg.



Bei Projekten zur Digitalisierung in den Geisteswissenschaften ist heute die Realisierung größerer, überregionaler und über das World Wide Web abfragbarer Lösungen erforderlich. Die Beiträge dieses Bandes leisten einen interdisziplinären Beitrag zur erforderlichen Standardisierung dieser Angebote, die erst den dringend notwendigen Austausch erleichtern und die gemeinsame Nutzung strukturierter Daten ermöglichen kann.

ISBN-10 3-937816-27-5

ISBN-13 978-3-937816-27-2 ISSN 0436-6638

Hamburg University Press
Verlag der Staats- und Universitätsbibliothek Hamburg