

Codices Electronici Ecclesiae Coloniensis

Historische Quellen im Internet bereitzustellen, hat mittlerweile durchaus Tradition. Der Verfasser kann auf das von ihm selbst schon Mitte der Neunziger Jahre angestossene Projekt zur Digitalisierung eines umfangreichen Archivbestandes verweisen.¹ Während in den späten Achtzigern und frühen Neunzigern Überlegungen zur digitalen Bereitstellung historischen Quellenmaterials jedoch überwiegend aus den historisch²-philologischen³ Disziplinen selbst kamen, haben seit Ende der Neunziger vor allem die Bibliotheken diese Idee unter dem Konzept der „Digitalisierung des Kulturellen Erbes“ aufgegriffen. Deutschland kann hier mit dem Strategiepapier des Ausschusses für Wissenschaftliche Bibliotheken und Informationssysteme für den Ausbau digitaler Ressourcen bis 2015,⁴ bzw. der daraus für den Bereich der „Kulturellen Überlieferung“ durch den zuständigen Unterausschuss abgeleiteten Aktionspläne,⁵ auf eines der anspruchsvollsten existierenden Förderprogramme verweisen, das sich die mittelfristige Gesamtdigitalisierung der mindestens bibliothekarischen Überlieferung nicht nur als abstraktes Ziel, sondern als konkretes, mit finanziellen Überlegungen unterfüttertes Aktionsprogramm vorgenommen hat.

I. Das Projekt

In dieser Bewegung weg von der Digitalisierung ausgezeichneter Einzelobjekte hin zur systematischen Konversion ganzer Sammlungen zeichnet der Verfasser an der Universität zu Köln für ein Projekt⁶ verantwortlich, in dem wir, in Zusammenarbeit mit der Diözesan- und Dombibliothek der Erzdiözese Köln, deren vollständigen Bestand an mittelalterlichen Codices digitalisiert haben, der im Wesentlichen immer noch den Zustand der mittelalterlichen Kathedralbibliothek wiedergibt, im Projekt ergänzt durch einige Stücke anderer kirchlicher Einrichtungen der Diözese.

Komplexe WWW-Ressourcen auf Papier zu beschreiben, ist stets etwas unbefriedigend; dem Interessenten wird der Besuch von <http://www.ceec.uni-koeln.de> daher dringend empfohlen. Zu den Hauptcharakteristika.⁷

- Alle (annähernd 420) Codices der Sammlung – ca. 150 000 Seiten – wurden digitalisiert.
- Sie sind über das WWW ohne jede Einschränkung verfügbar und zwar in Auflösungen bis zu ca. 4000×3000 Pixel für normale Formate und noch sehr viel höheren Auflösung für einige sehr großformatige Codices. Diese Auflösung genügt, um die meisten paläographischen Standardfragen zu beantworten; in der Regel ist noch feststellbar, in welcher Richtung das Schreibwerkzeug bewegt wurde.
- Die Bilder der einzelnen Seiten werden mit Bedienungselementen präsentiert, die es erlauben durch die Manuskripte zu blättern wie in einem Buch.
- Alle codicologischen Beschreibungen der Manuskripte seit dem 18. Jahrhundert wurden in eine Datenbank übertragen, die diese Beschreibungen für die Internetpräsentation verwaltet. Auf dieser Basis sind einige Suchwerkzeuge für die Beschreibungen definiert – allerdings liegt hierauf kein besonderes Gewicht, da die Anzahl der Codices ja noch gut überschaubar ist. Diese „Metadaten“, wie Bibliothekare solche Beschreibungen nennen, haben für manche Codices selbst den Charakter von eigenständigen Publikationen, im Umfang von bis zu fünfzig gedruckten Seiten. Sie werden jedoch vollkommen dynamisch verwaltet. Das bedeutet, dass alle Referenzen aus diesen Beschreibungen auf verschiedene Manuskriptseiten ad hoc zu Internetlinks auf die Digitalisate der einzelnen Seiten konvertiert werden – falls zu einem bestimmten Zeitpunkt die beschriebe-

nen Seiten bereits digitalisiert worden sind. Es bedeutet darüber hinaus aber vor allem, dass sogar die mehrseitige Beschreibung eines Manuskripts auf dem Bildschirm, die für den Benutzer wie die direkte Transkription aus einem gedruckten Buch aussieht, in Wirklichkeit ein Resultat einer Datenbankabfrage ist, die eine spezifische intellektuelle „Ansicht“ der Forschungsdiskussion über dieses Manuskript darstellt. Eine dieser Ansichten wird für Anfänger ohne besondere Kenntnisse von Manuskriptstudien bereitgestellt: In dieser Ansicht wird für jedes Manuskript eine vereinfachte Zusammenfassung aller gespeicherten Beschreibungen präsentiert. Nehmen wir an, in einer der codicologischen Studien – „a“ – wurde eine Aussage zur Entstehungszeit des Manuskripts gemacht, diese Studie sieht den Entstehungsort jedoch als unbekannt an, während eine andere Studie – „b“ – beides angibt, Entstehungszeit und Entstehungsort. In diesem Fall würde in der jetzt besprochenen „Laienansicht“ auf alle Fälle der Entstehungsort der Beschreibung „b“ ausgewiesen; ob die Entstehungszeit aus „a“ oder „b“ entnommen wird, wird auf der Basis einer „Präferenzordnung“ entschieden, die den relativen Wert der einzelnen Studien wiedergibt: entweder als systemweite Präferenzordnung oder als Teil des persönlichen Profils eines registrierten Benutzers. Gleichzeitig existiert aber auch noch eine andere „Ansicht“, die sich an den fortgeschrittenen Benutzer wendet, der sich der Widersprüchlichkeit von Aussagen in der Forschung bewusst ist: Ihm oder ihr würden die Aussagen aller vorliegenden Beschreibungen zu den einzelnen Punkten angeboten. In unserem Beispiel würde die Entstehungszeit deshalb zweimal wiedergegeben werden (geordnet wieder nach der anzuwendenden Präferenzordnung).⁸

- Die Codices werden innerhalb einer webbasierten Arbeitsumgebung⁹ präsentiert, die auch digitalisierte Versionen der wichtigsten Literatur enthält, die für die wissenschaftliche Arbeit mit den Texten nötig ist, so weit diese ohne Copyright-Beschränkungen verfügbar ist oder die Verfasser dieser Verwendung ihrer Werke explizit zugestimmt haben.
- Die Arbeitsumgebung, in der die Codices präsentiert werden, enthält darüber hinaus Werkzeuge, die die Durchführung typischer codicologischer/paläographischer Aufgaben erleichtern. Der Benutzer kann beispielsweise ein Werkzeug herunterladen, um für einzelne oder mehrere Zeichen ein Repertoire der grafischen Formen, die in einem bestimmten Codex oder einer Gruppe von solchen verwendet werden, zu erstellen.

II. Nutzung

Nur wenn vollständige Sammlungen, keine einzelnen Objekte, solcherart verfügbar gemacht werden, so glauben wir, kann diese Form der Quellenwiedergabe ein Standardmedium werden, das von Forschenden wirklich konsultiert wird. Wieweit dies geschieht zu quantifizieren ist schwierig. Der größte Teil des Materials besteht aus mittelalterlichen Handschriften, die mangels Bildschmuck keine unmittelbare ästhetische Anziehungskraft haben und häufig nicht besonders leicht zu lesen sind. Sie enthalten Texte in lateinischer Sprache, nicht unbedingt in besonders einfachem Latein. Sie diskutieren häufig diffizile Punkte mittelalterlicher kanonistischer Rechtsvorschriften und ähnliche Themenfelder. Mit anderen Worten: Man sollte annehmen, dass das Interesse der Laien sehr gering ist. In der Tat spricht das System explizit qualifizierte akademische Forscher an und macht dem interessierten Laien kaum Zugeständnisse. Trotz alledem wird das Material häufig betrachtet. Ob dies geschieht, weil es fachliches Interesse weckt oder – trotz all des Gesagten – dennoch das Interesse von Laien anzieht, kann nicht endgültig entschieden werden.

Immerhin haben wir im Rahmen einer im Auftrag der Deutschen Forschungsgemeinschaft finanzierten Studie¹⁰ über die Nutzung der digitalen WWW-Angebote für den Bereich des kulturellen Erbes insgesamt durch die Forschung einige interessante Beobachtungen gemacht. Dazu haben wir angenommen, dass ernsthaftes Interesse an dem Material nur besteht, wenn jemand mindestens fünf Manuskriptseiten jeweils mindestens 60 Sekunden lang betrachtet und keine Unterbrechung in der Benutzung des Angebotes von mehr als 20 Minuten eintritt. Bei diesen extrem konservativen Annahmen erhalten wir als Ergebnis immer noch, dass im Tagesdurchschnitt (wobei wir von 365 Tagen ausgehen) drei bis vier Interessenten oder Interessentinnen das Material interessant genug finden, um nach diesem Kriterium eine Stunde oder länger damit zu verbringen – und wenn sie dies einmal tun, innerhalb des Jahres meist auch mehrmals zu gleich langen „Sitzungen“ wiederkommen. Ob es sich dabei um „wissenschaftliche“ Interessenten handelt oder vielleicht um den Beleg, dass der beklagte Mangel des breiteren Publikums an den Schätzen traditioneller Bildung daran liegen könnte, dass diese Schätze – aus durchaus unterstützenswerten konservatorischen Überlegungen – für dieses breitere Publikum normalerweise einfach nicht zugänglich sind, überlasse ich Ihnen.

Wir glauben, dass dieser Erfolg damit zusammenhängt, dass wir einerseits bewusst versucht haben, eine hinreichend große kritische Masse von Material bereitzustellen, die es erlaubt, auch Dinge zu finden, von denen man vor dem Besuch des Servers nichts wusste – was bei den Angeboten „schöner und berühmter“ Stücke eben nicht zutrifft. Wir glauben auch, dass dieser Erfolg damit zusammenhängt, dass wir ganz gezielt versucht haben, Qualitäten bereitzustellen, die für die große Mehrzahl der forschungsrelevanten Fragen ausreicht. Nein, wir glauben nicht, dass die Betrachtung eines Digitalisates den Kontakt mit dem Original ersetzen kann. Wir glauben daher, dass dieses Gegensatzpaar die Realität der Forschung verkennt. Der Forscher und die Forscherin, die sich intensiv mit einem bekannten Stück beschäftigen wollen, werden natürlich nach wie vor daran interessiert sein, es im Original einzusehen. Diejenigen Kolleginnen und Kollegen, die einen Hinweis in einer von vielen Fußnoten vorbereiten, werden dies aber kaum tun können und haben, nachdem das Material so gut wie vollständig unedierte ist, auf diese Hinweise eben bisher verzichtet, selbst wenn sich aus der Vorbereitung dieses Hinweises vielleicht eine wichtige weiterführende Erkenntnis ergeben hätte. Dazu aber braucht man keine „schönen Stücke“ – wohl aber verlässliche Qualität.

Der Großteil des Materials ist deshalb in einer Auflösung gespeichert, die ca. 35–48 MB von Daten pro Manuskriptseite benötigt; für einige großformatige Codices verwahren wir in Langzeitspeicherung etwa 128 MB pro Seite. Dies beläuft sich auf ca. 3,5–4,5 TB (Terabyte) Datenmaterials oder, intuitiver, einem Stapel von ca. 6000–8000 CD-ROMs. Das Projekt stellt also wenig bescheidene Qualitätsanforderungen. Wir möchten jedoch hervorheben, dass, mit gutem Projektmanagement und Transparenz des zu Grunde liegenden technischen Prozesses, diese Art von Projekten mittlerweile „routinemäßig“ durchgeführt werden kann: Die Kosten der Digitalisierung pro Seite in der angezeigten Qualität lagen bei knapp über 1 Euro pro Seite und in Folgeprojekten lassen sich diese Kosten nochmals beträchtlich senken. (Falls – und nur falls – gutes Projektmanagement eingesetzt wird.)

III. Finden ohne Umwege

Ein weiterer beträchtlicher und bewusster Unterschied zwischen diesem Kölner Projekt und der großen Mehrzahl von Kulturerbe-Servern ist, dass es sich selbst in einem bestimmten Sinn als weniger wichtig und seinen Server als transparent betrachtet. Falls jemand in den Handapparat des CEEC-Servers, die Sammlung digitalisierter Literatur über Manuskripte, schaut, wird er oder sie unter

der Literatur zum Codex 124 einen Aufsatz von Gerhard Schmitz finden.¹¹ Wenn man diesen Text betrachtet, wird man bemerken, dass einige der traditionellen Referenzen zu einzelnen Seiten des Codex, die dort vorkommende Merkmale diskutieren (z. B. „fol. 25v, 26r“), als Links fungieren, die den Benutzer direkt zu den Seiten des Manuskripts bringen, die der Text von Gerhard Schmitz an dieser Stelle diskutiert.

Diese modernisierte Version einer Fußnote, die versucht, eine direkte Verknüpfung zwischen dem wissenschaftlichen Diskurs über die Quelle und dem Quellenbefund, auf den der Diskurs sich stützt, herzustellen, ist noch immer nicht alltäglich, jedoch insofern nichts Neues, als sie in der einschlägigen Diskussion über moderne Formen elektronischer Textdarstellung seit Jahren, eigentlich Jahrzehnten, gefordert wird. Die Kölner Manuskriptbibliothek zeichnet sich jedoch in einem sehr wichtigen Punkt aus: Sie ist wahrscheinlich die einzig bisher existierende, die explizit und systematisch einen Mechanismus zur Herstellung dieser Art von Links anbietet, der von jedem Verfasser eines externen Dokuments – auch ohne dass dies den Verwaltern der Manuskriptbibliothek bewusst ist – genutzt werden kann.¹²

Der Hintergrund für dieses Vorgehen muss etwas detaillierter erklärt werden. Digitale Bibliotheken, wie alle anderen neuen Anwendungen seit Erfindung der Datentechnik, versuchen zunächst stets, ihr Angebot möglichst exakt nach einem den prospektiven Benutzern bereits wohl bekannten Ablauf zu gestalten, um die Umstellung zu erleichtern. Mit anderen Worten, digitale Informationssysteme versuchen zunächst stets eine Metapher eines nicht digitalen Informationssystems zu verwenden, in unserem Fall die der traditionellen Bibliothek. Diese Metapher geht von einer Bibliothek aus, die an einem spezifischen geographischen Punkt existiert, wohin man mit verschiedenen Transportmitteln reisen kann; ist man an ihrem Standort eingetroffen, geht man durch eine Eingangshalle, bewegt sich zu den Katalogen dieser speziellen Bibliothek, sucht darin nach dem gewünschten Werk, entscheidet auf der Basis der knappen Information, die auf einer Katalogkarte Platz findet, ob man es braucht, bestellt es bei der Ausleihe, wartet und kann es schließlich im Lesesaal der Bibliothek einsehen. Handelt es sich um ein Manuskript, das wegen seines Wertes nicht transportiert werden kann, muss jede Kontrolle eines Zitats, auch wenn dieses Zitat die Signatur enthält, diesen ganzen Prozess, einschließlich der vorgeschalteten Fahrplanstudien, wiederholen.

Bei der Beschreibung dieser Metapher werden die Bibliothek und ihre konstituierenden Teile viel öfter genannt, als der gesuchte Text, obgleich die meisten Leser an der Bibliothek als solcher völlig uninteressiert sind und sie nur besuchen, weil sie dazu gezwungen sind, um den Codex ihres Interesses zu konsultieren. Trotzdem wurde dieses Modell des Informationszuganges, wo sehr viel von der Bibliothek und sehr wenig vom Text die Rede ist, detailgenau in die digitale Welt übertragen. Auch wenn man eine digitale Bibliothek nutzen möchte, muss man deren URL kennen, die Struktur ihrer Homepage verstehen, die Eigenheiten ihres OPAC durchschauen, Buch oder Manuskript nach dem Zitat in diesem finden und den Text schließlich innerhalb der spezifischen angebotenen Arbeitsumgebung betrachten. Das geht schneller als in der realen Welt – die Schritte aber sind immer noch dieselben.

Unseres Erachtens sind die meisten davon prinzipiell völlig überflüssig: Wir gehen davon aus, dass der Benutzer ein spezifisches Manuskript ansprechen möchte, ohne damit belästigt zu werden, wo es aufbewahrt wird, sei es physisch oder virtuell. Dies sollte im Prinzip in digitalen Informationssystemen ganz einfach möglich sein, da die Geisteswissenschaften seit Generationen ein System von Identifikationen für Manuskripte verwenden, die für alle praktischen Zwecke „Permanent Universal Resource

Identifiers“ im Sinne der Informatik sind, also unveränderliche eindeutige Referenzen. Technisch können diese leicht verwandelt werden zu „Permanent Universal Resource Locators“, mit deren Hilfe man sie direkt auf den Bildschirm bringen kann. Im Falle des Aufsatzes von Gerhard Schmitz bezieht sich das Zitat auf „Folio 25v von Codex 124 der ‚Kölner Diözesan- und Dombibliothek““. Wenn wir dies umformen zu „‚Kölner Diözesan- und Dombibliothek‘ Codex 124; Folio 25v“ haben wir nun eine etwas besser geordnete Form; falls wir den Namen dieser Bibliothek durch die Sigle ersetzen, unter der sie im deutschen Bibliothekssystem bekannt ist, bekommen wir eine weniger lesbare, aber immer noch verständliche Version: „kn28-Codex 124; Folio 25v“; formalisieren wir die Referenz zu dem Codex etwas weiter, bekommen wir „kn28-0124_;Folio 25v“ und machen wir das Gleiche mit der Folioreferenz, bekommen wir „kn28-0124_25v“.¹³

Dies sieht technisch aus, und ein Werkzeug, um die Namen der Bibliotheken in ihre Siglen zu konvertieren, wäre hilfreich; aber zumindest für einen Manuskript-Bibliothekar wäre die Referenz dennoch unmittelbar verständlich. Und wenn wir sie in die Adresszeile eines Browsers mit der hinzugefügten Verpackung „http://www.ceec.uni-koeln.de/ceec-cgi/kleioc/0010KICEEC/exec/pagemed/%22|kn28-0124_25v%22“ (die wiederum sehr leicht durch einen Nachschlagemechanismus bereitgestellt werden könnte) eintragen, finden wir uns direkt auf der angesprochenen Manuskriptseite wieder – umgeben von passenden Buttons, um in dem Codex blättern und auf alle anderen Werkzeuge der Arbeitsumgebung der digitalen CEEC-Bibliothek zugreifen zu können. Mit anderen Worten: Wir können exakt die gleichen Dinge tun, die wir tun könnten, wenn wir die Bibliothek durch die Türe – die Homepage – betreten hätten.

Manfred Thaller

Zum Autor

Prof. Dr. MANFRED THALLER, Professor für Historisch Kulturwissenschaftliche Informationsverarbeitung an der Universität Köln

Kontakt: manfred.thaller@uni-koeln.de

Website des Projekts: <http://www.ceec.uni-koeln.de>

Anmerkungen

- 1 Manfred Thaller: Digitale Archive: Technik und Methode, in: Hans-Heinrich Ebeling und Manfred Thaller (Hg.): Digitale Archive. Die Erschließung und Digitalisierung des Stadtarchivs Duderstadt, Göttingen 1999, 125-162.
- 2 Für ein ursprünglich auf CD-ROM Basis konzipiertes elektronisches Editionsprojekt vgl. die jetzt als Web-Edition verbreiteten Fontes Civitatis Ratisponensis unter <http://bhgw20.kfunigraz.ac.at/>
- 3 Kevin S. Kiernan: Beowulf and the Beowulf Manuscript, Chicago 1997, geht trotz des späten Erscheinungsdatums auf ein 1987 begonnenes Projekt zurück.
- 4 http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/positionspapier.pdf
- 5 http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/pos_papier_kulturelle_ueberlieferung_0511.pdf
- 6 Vgl. Codices Electronici Ecclesiae Coloniensis. Eine mittelalterliche Cathedralbibliothek in digitaler Form, hg. von Manfred Thaller (Fundus Beih. 1, Göttingen 2001).
- 7 Die folgende Darstellung entspricht weitgehend der in Manfred Thaller: Reproduktion, Erschließung, Edition, Interpretation: Ihre Beziehungen in einer digitalen Welt (in: Brigitte Merta, Andrea Sommerlechner und Herwig Weigl [Hg.]: Vom Nutzen des Edierens, Wien 2005 [= MIÖG Erg. Band 47], 205-227) gegebenen.
- 8 Zu den dabei verwendeten technischen Konzepten siehe Manfred Thaller, Texts, Databases, Kleio: a Note on the Architecture of Computer Systems for the Humanities, in: Augmenting Comprehension. Digital Tools and

- the History of Ideas, hg. von Dino Buzzetti–Giuliano Pancaldi–Harold Short (Office for Humanities Communication Publication 17, London 2004), 49–76.
- 9 Die innerhalb des Projektes angebotenen Werkzeuge für die paläographische Arbeit von Patrick Sahle finden sich auf <http://www.ceec.uni-koeln.de> unter dem Reiter „Werkzeuge“. Als Beispiel eines in der Literatur voll dokumentierten Werkzeuges für den Umgang mit digitalisierten Manuskripten sei auf Armin Lange, Computer Aided Text-Reconstruction and Transcription. CATT-Manual (Tübingen 1993), verwiesen.
 - 10 http://www.dfg.de/forschungsfoerderung/wissenschaftliche_infrastruktur/lis/download/retro_digitalisierung_eval_050406.pdf
 - 11 Nach der Druckausgabe: Gerhard Schmitz, Die Vier-Bücher-Sammlung des Cod. Köln, Diözesan- und Dombibliothek 124. Zur kirchenrechtlichen Kenntnis im 10. Jahrhundert, in: *Ex Ipsis Rerum Documentis. Beiträge zur Mediävistik. Festschrift für Harald Zimmermann zum 65. Geburtstag*, hg. von Klaus Herbers–Hans Henning Kortüm–Carlo Servatius (Sigmaringen 1991), 233–255.
 - 12 Zur grundsätzlichen Nutzung von hypertextorientierten Strukturen als editorischem Werkzeug innerhalb einer Edition (also nicht, wie bei uns, zwischen einer Quellenrepräsentation und dem diese Quelle interpretierenden Text) vgl. Federico Pellizi, Hypertext as a Critical Discourse: From Representation to Pragmatics, Scholarly editing in the digital age, in: *New Media and the Humanities: Research and Applications. Proceedings of the First Seminar Computers, Literature and Philology*, Edinburgh, 7–9 September 1998, hg. von Domenico Fiormonte–Jonathan Usher (Oxford 2001), 57–67.
 - 13 Zu diesem Mechanismus aus der Sicht digitaler Bibliotheken Manfred Thaller, From the Digitized to the Digital Library, in: *D-Lib Magazine* (February 2001), <http://www.dlib.org/dlib/february01/thaller/02thaller.html>

Auszug aus:

**Jahrbuch der historischen Forschung
in der Bundesrepublik Deutschland 2006**

Herausgegeben von der
Arbeitsgemeinschaft historischer Forschungseinrichtungen
in der Bundesrepublik Deutschland e.V.

© 2007 Oldenbourg Wissenschaftsverlag GmbH, München
oldenbourg.de

Das Werk einschließlich aller Abbildungen ist urheberrechtlich geschützt.
Jede Verwertung außerhalb der Grenzen des Urheberrechtsgesetzes ist ohne
Zustimmung des Verlages unzulässig und strafbar. Dies gilt insbesondere für
Vervielfältigungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung
und Bearbeitung in elektronischen Systemen.