

# Die Handschriftenbibliothek des Kölner Doms im Internet

von MANFRED THALLER  
*Universität zu Köln*

Über Internetpräsentationen zu schreiben, ist nicht wirklich schwierig; bis zu einem gewissen Grade scheint es aber überflüssig, da ein gut realisiertes Projekt sich eigentlich selbst beschreiben sollte. Vor allem anderen also nochmals die schon im Vorwort zu diesem Bändchen zitierte URL: <http://www.ceec.uni-koeln.de>.

Zusammenfassend: Das hier beschriebene Projekt ist angetreten, den "Gesamtbestand" an mittelalterlichen Handschriften an der Diözesan- und Dombibliothek im Internet in einer Form zugänglich zu machen, die den klassischen Benutzern dieses Bestandes den Zugriff darauf erlaubt und gleichzeitig die Möglichkeiten zu seiner Nutzung in Lehre und Forschung verbessert. Wie dies realisiert wurde, kann am besten im direkten Zugriff erprobt werden. Ein Wort der Warnung: Das Projekt geht kompromisslos davon aus, dass die Sammlungen in einer Qualität angeboten werden müssen, die für die praktische Arbeit mit den Manuskripten ausreicht. Dies, nicht die Erreichbarkeit von jedem Rechner aus, war oberste Richtschnur. Die übertragenen Datenmengen sind groß und setzen schnelle Leitungen voraus.

Diese Darstellung versucht also nicht, zu beschreiben, was den Besucher des Manuskriptservers erwartet. Sie beschreibt Überlegungen, die aus dem Besuch der Webseite so nicht unmittelbar hervorgehen.

## 1. Qualität

Was bedeutet es, Handschriften "kompromisslos in einer Auflösung bereitzustellen, die für die praktische Arbeit mit den Manuskripten ausreicht"?

In einem anderen Kontext<sup>1</sup> hat der Verfasser vorgeschlagen, die Qualität digitaler Objekte dadurch zu charakterisieren, wieweit sie innerhalb eines bestimmten An-

---

<sup>1</sup> 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, hier: 134 - 139.

wendungsbereiches einen funktionalen Ersatz für das Original darstellten, wieweit, beim Zugriff auf das Digitalisat, der Zugriff auf das Original also überflüssig würde. Hier haben wir vier Ebenen vorgeschlagen, deren Bezeichnungen aus den Eigenschaften von anderen Manuskriptservern abgeleitet wurden. Wir rekapitulieren kurz:

Ein digitales Objekt wird danach *illustrierend* genannt, wenn seine Qualität hinreicht, um eine qualifizierte Entscheidung darüber zu treffen, ob sich der Aufwand lohnt, auf das Original zuzugreifen. Dieses Niveau wird vor allem in Museumssystemen für Sammlungen verwendet, bei denen der Gesamteindruck eines Originals nach wie vor weit über das hinausgeht, was durch eine Wiedergabe auf dem Bildschirm erreicht werden kann (wie er innerhalb der Forschung heute zur Verfügung steht).

Ein digitales Objekt wird *lesbar* genannt, wenn seine Qualität erlaubt, auf all die Information zuzugreifen, die der Erzeuger des ursprünglichen Objekts an den Benutzer übermitteln wollte. (Die Seiten eines digitalisierten Buches müssen klar und ohne Belastung der Augen lesbar sein. Es ist nicht notwendig, die handschriftlichen Notizen lesbar zu halten, die von Generationen von Studenten in der Panik knapp vor den Abschlussprüfungen hinterlassen wurden.)

Ein digitales Objekt wird *paläographisch* genannt, wenn die gewählte Qualität all jene Informationen wiedergibt, die vom Original durch das unbewaffnete Auge abgelesen werden können. (In mittelalterlichen Codices ist es wichtig, den Text zu lesen; es ist jedoch auch wichtig, festzustellen, ob eine Änderung in der Art, in der die Feder gehalten wurde, in den Buchstaben noch erkennbar ist und einen Hinweis auf einen Wechsel des Schreibers liefert.)

Ein digitales Objekt wird, schließlich, *enhanceable* genannt, wenn der Zugriff auf die digitale Version mehr an Information bereitstellt, als durch das unbewaffnete Auge enthüllt wird. (Techniken des Image enhancement können, beispielsweise, Rasuren wieder lesbar machen.)

Unserer Ansicht nach ist jede Qualität unterhalb des Niveaus der Lesbarkeit in Bibliothekssystemen fast immer sinnlos; und selbst in solchen Fällen, in denen es Gründe für die Verwendung illustrierender Objekte gibt, sollte man die Definition "ausreichend um eine informierte Entscheidung darüber zu treffen, ob sich der Zugriff auf das Original lohnt" im Auge behalten. Als Konsequenz davon scheint es

uns z.B. wenig sinnvoll, die in Museumskontexten sehr nützlichen Thumbnails unverändert auf Bibliothekssysteme zu übertragen, wo sie dann eine Aneinanderreihung von zahlreichen meist so gut wie gar nicht mehr unterscheidbaren Seiten bilden.

Wenn wir dieses Modell, die angemessene Qualität der Bilder in einem digitalen System von einem Konzept der Benutzeranforderungen abzuleiten - anstatt von absoluten Auflösungszahlen - ernst nehmen, folgt daraus, dass die Spezifikation einer digitalen Bibliothek eine Reihe von Annahmen explizit machen muss: Welche Qualität der Hardware ihrer Benutzer unterstützt werden soll, welche Absichten sie nach Ansicht der Bibliothek bei Benutzung ihrer Bestände haben und die konkreten Lösungen, die aus diesen Vorgaben angeleitet werden.

Im Falle unserer Manuskriptbibliothek stellen wir folgendes Modell zur Diskussion.

Annahmen:

- a) Professionelle Arbeit mit Manuskripten am Bildschirm setzt als absolutes Minimum Bildschirme mit einer Auflösung von 1024 x 768 Bildpunkten voraus. Es werden daher keinerlei Vorkehrungen getroffen, *analytische* Arbeit auf niedrigeren Auflösungen zu unterstützen.
- b) 1024 x 768 definiert jedoch wirklich das absolute *Minimum*. Eine Auflösung von 1200 x 1024 wird wärmstens empfohlen, wenn mehr als die reine Lektüre gewährleistet werden soll. Vorkehrungen dafür sind daher zu treffen.
- c) 1600 x 1200 wäre eigentlich sinnvoll. Die derzeitige Ausstattung der Forschungseinrichtungen bietet diese Auflösung jedoch so selten an, dass noch keine konkrete Unterstützung dafür angeboten wird.

(Es sei am Rande angemerkt, dass u.E. Manuskriptarbeit an Bildschirmen mit einer Größe von weniger als 19 Zoll sinnlos ist. Da die physikalische Größe der Bildschirme - zum Unterschied von ihrer Auflösung - jedoch keine Auswirkungen auf die anzubietenden Bildqualitäten hat, verfolgen wir dieses Argument hier nicht weiter.)

Auf der Grundlage dieser Annahmen wurden die folgenden Auflösungen für die praktische Arbeit definiert:

Die niedrigste Auflösung, ist definiert durch *visuelle Zusammenfassungen*, die dem

Benutzer im rechten Bildschirmteil begegnen, wenn er die Manuskripte durchblättert, bzw., wenn er Sie von einer anderen Seite mit dem Symbol "bloßes Auge" anspricht. Sie liegt *über* der Auflösung der klassischen Thumbnails, so dass eine informierte Entscheidung darüber getroffen werden kann, ob diese Seite nützlich ist, eignet sich auf den angesprochenen Geräten jedoch nicht zur fortlaufenden Lektüre. (Kann ggf. aber noch als Notbehelf auf Geräten mit einer Auflösung von 800 mal 600 dienen.)

Für die reguläre Arbeit sind zwei höhere Auflösungen, die grundsätzlich bildschirmfüllend dargestellt werden, vorgesehen: *Arbeitskopien*, erreichbar über das Symbol "Brille", stehen in einer Auflösung zur Verfügung, die es erlaubt das Blatt auf einem 1024 x 768 Schirm ohne horizontales Scrollen flüssig zu lesen. Diese Auflösung bewahrt die meisten optischen Charakteristika des Originals, nimmt dafür jedoch eine gelegentliche Kontrastchwäche in Kauf. *Optimierte Arbeitskopien*, erreichbar durch eine "Lupe", werden in einer höheren Auflösung angeboten, die es erlaubt, die horizontale Auflösung eines 1200 x 1024 Schirms voll auszunutzen. (Sie ist in der Praxis jedoch so gewählt, dass der eigentliche Schriftblock der Seiten auch in die horizontale Dimension eines 1024 x 768 Schirmes passen wird. Diese Seiten werden auf mechanischem Wege optimiert, d.h., einige Kontrastverschärfungstechniken u.ä. werden auf die Seiten angewandt um die Lesbarkeit der Seiten zu verbessern, soweit dies möglich ist, ohne die anzuwendenden Bildverbesserungsverfahren auf eine (angesichts der Seitenzahlen zu zeitintensive) individuelle Analyse der Charakteristika der Einzelseiten zu stützen. Der Preis für diese Qualitätssteigerung liegt in Verzerrungen vor allem der Farben.

Für vergleichsweise seltene Fälle detaillierter fachlicher Arbeit, vor allem im Bereich der Paläographie, wird eine ziemlich hohe, durch ein "Mikroskop" ansprechbare, *Maximalauflösung* nahe an 4491 x 3480 Pixeln angeboten. Wir sind stolz darauf, damit eine Auflösung in den Bereich digitaler Bibliotheken einzubringen, wie sie bisher nur in Projekten Anwendung fand, die sich auf CD-ROM Basis mit ausgewählten Einzelstücken beschäftigen, wie etwa dem "Beowulfprojekt". In vereinzelten Fällen, in denen visuelle Charakteristika (insbesondere Goldauflage) durch den standardmäßig eingesetzten Digitalisierungsprozess nicht hinreichend wiedergegeben werden, werden für einzelne Seiten schließlich noch *Detailaufnahmen*, mit dem Symbol "Schiere" (für "Ausschnitt") wiedergegeben, die dann selbst wieder in allen vier eben beschriebenen Grundaufösungen bereit stehen.

## 2. Erschliessung

Innerhalb des CEEC Projektes steht derzeit erst eine vergleichsweise rudimentäre Ausstattung von Zugriffsinstrumenten im Sinne von Suchmaschinen zur Verfügung. Dafür gibt es zwei Gründe.

In einer Sammlung von etwas unter 400 Codices scheint ein elaborierter Metadatenzugang auf der Ebene etwa eines OPAC nicht allzu relevant: Hier stehen essenziell vorsortierte Listen der Codices nach den wichtigsten üblichen Kriterien zur Verfügung, die unter dem Punkt "Stöbern" angesprochen werden können.

Suchmaschinen, als Volltextsuchen innerhalb der sehr ausführlichen Beschreibungen einzelner Handschriften auf der Basis bestehender codicologischer Verzeichnisse realisiert, sind über den Punkt "Suchen" zugänglich. Sie sind allerdings ausdrücklich als "vorläufig" zu bezeichnen. Im Wesentlichen erlauben sie einen Volltextzugriff auf jene Teile der beschreibenden Daten, die entweder Quellzitate enthalten - vor allem incipit und explicit der Handschriften oder Handschriftenteile - oder natürlich sprachliche längere Beschreibungen von Manuskripten, Manuskriptteilen oder Buchillustrationen darstellen.

Dass hier vorerst kein größerer Aufwand getrieben wurde, liegt vor allem daran, dass hier noch der Kontakt zu den inhaltlichen Benutzern gesucht wird um die wirklichen Bedürfnisse zu klären. Die Beschreibungen der Manuskripte gehen über das in Bibliothekssystemen und Handschriftendatenbanken übliche um Größenordnungen hinaus: Dementsprechend wären theoretisch extrem elaborierte Suchmasken, die es erlauben Handschriften eines bestimmten Formats, die in aufeinander folgenden Abschnitten zwei incipits mit bestimmten Suchworten enthalten und vor dem Jahre  $x$  auf einem bestimmten Beschreibstoff geschrieben wurden, realisierbar. Bevor geklärt ist, was davon *wirklich* benötigt wird, nehmen wir von derlei Dingen aber Abstand: Das Forschungsinteresse, das durch obige Abfrage unterstützt würde, ist ein wenig schwierig vorstellbar - und es besteht der Verdacht, dass, wer eine Abfrage dieser Art innerhalb eines Bestandes von 400 Codices inhaltlich sinnvoll formulieren könnte, das Material so gut kennen muss, dass er die Antwort vermutlich ebenfalls bereits kennt. Dem anzweifelbaren Nutzen dieser Art von Suchinstrument stünde der abschreckende Charakter der dafür notwendigen komplexen Abfragemasken gegenüber, der unseres Erachtens der gewünschten nahtlosen Integration der Benutzung digitaler Manuskripte in den Arbeitsalltag gerade auch jener For-

scherrinnen und Forscher, die *keine* besondere technische Expertise besitzen, eher hinderlich als förderlich entgegenstünde. Und redundante Werkzeuge zu realisieren, nur weil dies technisch geht, scheint wenig ersprießlich.

Dennoch sind die bereitgestellten Metadaten extrem umfangreich: Kurz gesagt stehen für die Handschriften die Beschreibungen der gesamten über das Webinterface unter "Literatur" unter der Rubrik "Kataloge (im weitesten Sinne)" angeführten Titel als integrierte Beschreibungen des Materials zur Verfügung. Diese Kataloge wurden in Anlehnung an den MASTER Beschreibungsstandard als XML Dateien bereit gestellt; zu den Erfahrungen dazu berichtet Patrick Sahle in seinem Beitrag "Erfahrungen mit internationalen Standards bei der Handschriftenbeschreibung" in diesem Bändchen. In unserem Kontext ist ein Prinzip wichtig: Die "Katalogdaten", die von der hinter dem Manuskriptserver verwendeten Datenbank stehen, stehen frei formulierten Texten, wie sie aus klassischen Manuskriptbeschreibungen bekannt sind, wesentlich näher, als den streng auf eine Vielzahl von Feldern aufgeteilten Beschreibungen herkömmlicher Manuskriptdatenbanken. Sie werden trotzdem "als Datenbank" verwaltet, wie dies durch Ansätze möglich wird, die im Bereich sogenannter "XML Datenbanken" derzeit diskutiert werden.

Dieses Prinzip "ein Text kann auf Grund der in ihm enthaltenen impliziten Struktur so behandelt werden, als stellte er eine Datenbank dar" ist uns äußerst wichtig. Die im Laufe der Jahrzehnte gewachsenen Usancen der Beschreibung von Handschriften - und durchaus auch die Usancen zu deren drucktechnischer Darstellung - scheinen uns ein Kapital darzustellen, das nicht wegen falsch verstandener "Bedingungen" für den Einsatz von Datenbanksoftware vergeudet werden sollte.

Im Falle des CEEC Servers ist diese Verbindung zwischen "Texten" und "Datenbanken" keineswegs bereits zu hundert Prozent gelungen: Die Auswirkungen dieses Prinzips sind jedoch teilweise leicht zu übersehen und werden daher hier näher beschrieben.

Kurz gesagt: *Keiner* der Texte, die als Beschreibungen von Handschriften oder Handschriftenteilen auf dem Bildschirm auftauchen, stehen in genau dieser Form in der Datenbank. Dies gilt auch und insbesondere für die über "Stöbern" zugänglichen "Katalogisate", die durch die drei Icons mit unterschiedlich ausführlich gefüllten Seiten als "Katalogdaten Kurzanzeige", "Katalogdaten Normalanzeige" und "Katalogdaten ausführliche Anzeige" abgerufen werden können. Alle diese Manuskript-

beschreibungen - in einzelnen Fällen das Äquivalent von 20 bis 30 Druckseiten - werden dynamisch, als Ergebnis einer Anfrage an eine Datenbank, errechnet.

Dieser recht ehrgeizige Mechanismus funktioniert, wie gesagt, noch nicht vollständig: Wenn in den Katalogisaten Brüche zwischen Satzteilen auftauchen - oder Satzzeichen falsch positioniert sind - stellt dies keinen Tippfehler in einer einmal geschriebenen HTML Seite dar, sondern ein Programmproblem bei der dynamischen Generierung des fortlaufenden Textes aus der zu Grunde liegenden Datenbank.

Der Grund für den recht hohen technischen Aufwand, der hier getrieben wird, sei an zwei Beispielen erläutert, wobei für das bessere Verständnis des ersten auf den Codex Köln 83ii Bezug genommen wird. Die "ausführliche Anzeige" der Katalogdaten liefert hier folgende Angaben zur Entstehungszeit:

Entstehungszeit: um 805 (nach: Zensus: 620);  
8/9c (nach: Anderson);  
dated 798 and 805 (nach: Anderson);  
798 und 805 (nach: Glaube und Wissen: 136);  
a. 798 scriptus (nach: JafféWattenbach: 29)

Diese Angaben werden für diese Anzeige der Katalogdaten dynamisch erstellt. Das heißt:

- (a) Sämtliche für den Codex verfügbaren Angaben zur Entstehungszeit werden aus den einzelnen, getrennt voneinander verwalteten, Beschreibungen zusammen getragen.
- (b) "Anklickbare Links" zu den einzelnen bibliographischen Angaben werden ad hoc erstellt.
- (c) Die Ordnung der angezeigten "Forschungsmeinungen" folgt keinen besonderen Vorgaben.

Bei der "normalen" Anzeige der Katalogdaten wird für denselben Sachverhalt und denselben Codex nur mitgeteilt:

Entstehungszeit: 798 und 805

Hier geschieht Folgendes:

- (a) Sämtliche für den Codex verfügbaren Angaben zur Entstehungszeit werden aus den einzelnen, getrennt voneinander verwalteten, Beschreibungen zusammen getragen.
- (b) Sie werden mit einer Liste der "Autoritäten" der einzelnen Handschriftenbeschreibungen verglichen, die festlegt, welche relative "Qualität" die einzelnen Beschreibungen haben.
- (c) Nur die nach diesem Regelwerk "beste" wird dem Benutzer, ohne weitere Verbindungen zur Quelle, mitgeteilt.

Derzeit werden diese beiden Katalogisatformen in dieser Weise angeboten um dem Benutzer der "Normalanzeige" zu jedem über den Codex bekannten Sachverhalt die mutmaßlich verlässlichste Forschungsmeinung mitzuteilen. Im ausführlichen Katalogisat wird dagegen versucht, die gesamte Tradition der Forschungsmeinungen über den betreffenden Sachverhalt zusammen zu stellen.

In weiterer Folge werden Mechanismen dieser Art wohl dazu verwendet werden um, wohl auch in Verbindung mit Benutzerprofilen für einzelne Benutzer, den besonderen Bedürfnissen angepasste Beschreibungen generieren zu können. Für den allgemeinen Benutzer z.B. eine der jetzigen Normalanzeige entsprechende Form; für den Teilnehmer an einem Proseminar ausschließlich die in einer bestimmten Beschreibung enthaltenen Angaben; für an eben diesem Codex arbeitende Spezialisten eine Darstellung, die in allen Fällen, in denen von ihnen eigene Entscheidungen getroffen wurden, eben diese, in allen anderen Fällen alle existierenden Forschungsmeinungen in chronologischer Folge anzeigt.

Ein weiterer Effekt dieser absolut dynamischen Verwaltung aller Beschreibungen liegt darin, dass sich der Server der gerade vorliegenden Informationsmenge weitgehend von selbst anpasst. Das heißt z.B., dass zusätzliche Beschreibungen oder Hinweise von spezialisierten Forschern ohne Umstellungen in den laufenden Server eingebracht werden können. Dies war aber vor allem auch einer der wesentlichen Gründe dass der Manuskriptserver, sehr ungewöhnlich für digitale Bibliothekssysteme dieser Art, bereits etwa vier Monate nach Projektbeginn den Probebetrieb aufnahm und nach neun Projektmonaten auch offiziell freigegeben werden konnte.

Da *alle* Verbindungen, also auch die einigen hunderttausend, die zwischen Beschreibungselementen in den Katalogisaten und den digitalisierten Seiten selbst an-

geboden werden, zwischen Metadaten und den durch sie dokumentierten Handschriftenseiten dynamisch, während der Generierung der angeforderten Katalogisate, errechnet werden, konnte zu einem sehr frühen Zeitpunkt ein erster Server angeboten werden, der zunächst nur für wenige Handschriften auf die digitalisierten Seiten zugriff. Da er jedoch bei jedem Zugriff "nachsieht", welche Manuskriptseiten bereits digitalisiert sind, und nur jene Links generiert, für die auch wirklich Verweisungsziele existieren, erweitert er sein Angebot ohne weiteres zutun, abgesehen vom Einspielen der Digitalisate, automatisch.

### **3. Kontext**

Bibliotheken bestehen nicht aus Bücherspeichern, aus denen die Werke durch einen Schalter an die auf der Straße stehenden Benutzer ausgegeben werden, sondern aus einer Arbeitsumgebung, in der ihm oder ihr die zur Benutzung notwendigen Mittel im Lesesaal zur Verfügung stehen. Dies muss natürlich auch bei digitalen Bibliotheken berücksichtigt werden; vor allem wenn Sie Objekte, wie die Handschriften der Diözesan- und Dombibliothek bereitstellen, zu deren Bearbeitung spezialisierte und nicht ohne weiteres verfügbare Literatur benötigt wird.

Leider zeigen die Copyrightregelungen hier klare Grenzen auf. Das Projekt war jedoch bemüht Literatur, die entweder außerhalb des Copyrights ist, oder für die wir die Genehmigung der Verfasser haben, in die Benutzungsoberfläche des Servers zu integrieren - ein Apparat, der über die Schalter "Historische Bibliothek" bzw. "Literatur" zugänglich ist. Je nach Alter und Aussehen des Materials wurden hier entweder durch OCR umgesetzte Texte oder die digitalisierten Seiten der Drucke bereitgestellt. In einigen Fällen gab es Punkte in der Tradition der Bibliothek, die ein detaillierteres Eingehen, über die vorliegende Literatur hinaus erforderten: Auf sie wird in den Beiträgen von Patrick Sahle und Torsten Schaßan näher eingegangen.

Im Allgemeinen ist dieser Aspekt der digitalen Bereitstellung der Handschriften der Diözesan- und Dombibliothek zu Köln aber sicher am transparentesten und weitgehend selbst erklärend, weshalb hier nicht weiter darauf eingegangen wird.

#### 4. Modularität

In keiner Weise selbst erklärend ist eine andere der konsequent durchgehaltenen Techniken bei der Erstellung des Handschriftenservers. In fast allen Stadien der Einführung der Informationsverarbeitung über die ersten rein technisch-ingenieurwissenschaftlichen hinaus, wurde beim Ersatz herkömmlicher Arbeitstechniken meist eine Metapher des zu ersetzenden Systems verwendet, die versuchte dies so gut als möglich abzubilden. Dies hat den unbestreitbaren Vorteil, dem Benutzer eines neuen digitalen Informationsdienstes den Überstieg von der herkömmlichen zur neuen Form der Informationsbeschaffung oder -verarbeitung so einfach als möglich zu machen; wir selbst haben eben die Metapher des Lesesaals verwendet um den Aufbau der Arbeitsumgebung für die digitalisierten Manuskripte zu beschreiben.

Doch: Metaphern können die weitere Entwicklung digitaler Informationssysteme auch behindern. Dies könnte in weiterer Folge gerade für Bibliotheks- und ähnliche Systeme im Bereich des Kulturerbes (Archive, Bibliotheken, Museen) gelten. Finden wir ein Literaturzitat in der Anmerkung eines Aufsatzes, so betraten wir bisher die lokale Bibliothek durch ihr Portal, konsultierten die Kataloge, forderten über die in ihnen enthaltene Information das zitierte Buch an um den weiteren Kontext der interessierenden Referenz kennen zu lernen. Dieser Prozess wurde durch die Zugänglichkeit von OPACs vom eigenen Schreibtisch aus erheblich beschleunigt; noch wesentlich erheblicher, wenn der jeweilige OPAC direkt mit Digitalisaten der fraglichen Literatur verbunden ist. Der grundsätzliche Zugangsweg "durch den Eingang in die Bibliothek, dort zum Katalog und von dort zum Buch" wird in der Mehrzahl existierender digitaler Bibliotheken aber nach wie vor exakt nachgebaut.

Im Falle des CEEC Servers wurde bewusst auch ein alternativer Weg beschritten, der am leichtesten zu erläutern ist, wenn der Leser innerhalb der Oberfläche den Schalter "Literatur" wählt und bei der auf den Codex 124 bezogenen Literatur den Aufsatz von Gerhard Schmitz: "Die Vier-Bücher-Sammlung des Cod. Köln, Diözesan- und Dombibl. 124 - Zur kirchenrechtlichen Kenntnis im 10. Jh." aktiviert. Wie der Benutzer unschwer feststellen kann, sind in diesem Aufsatz sämtliche Bezugnahmen auf den Codex durch Links auf die Digitalisate der zitierten Seiten dargestellt. Das scheint zunächst ein Ergebnis der Tatsache, dass dieser Aufsatz eben innerhalb des von uns postulierten "Lesesaals" der digitalisierten Handschriftenbibliothek liegt. Genau das ist aber nicht der Fall: Derselbe Aufsatz könnte völlig unab-

hängig von diesem Server exakt die gleiche Eigenschaft haben.

Grundsätzlich: Ist einem interessierten Benutzer oder einer interessierten Benutzerin das kanonische Zitat (d.h., das Zitat in der in den benutzenden Disziplinen üblichen Form) eines Codex der Diözesan- und Dombibliothek bekannt, z.B. Seite 82 recto von Codex 83ii, so reichen diese Angaben aus um, ohne Berücksichtigung der restlichen Struktur des Servers, eine direkte Verbindung mit eben dieser Seite herzustellen. Als willkommener Nebeneffekt dieses Bauprinzips ist die Wiedergabe der Handschriften im Internet damit auch ganz gezielt für die Einbindung in größere virtuelle Systeme vorbereitet. Ein derartiger Mechanismus, der die Metapher der einzelnen Sammlung und Bibliothek potentiell auflöst scheint uns für die intensivere Verbindung von digitalen Quellen und Darstellungen innerhalb der Beschäftigung mit dem kulturellen Erbe von ganz zentraler Bedeutung.

Wir beschreiben die hier verwendeten Konzepte daher in größerem Detail als andere Aspekte des Servers. Selbstverständlich sind derartige Überlegungen langfristig nur sinnvoll, wenn die angeführte Art der Zitate selbst langfristig gültig bleibt. Dies kann durch ein einzelnes Projekt natürlich nicht endgültig herbeigeführt werden; die dazu führenden Überlegungen müssen in einem Modell, das Nachhaltigkeit beansprucht, schon jetzt berücksichtigt werden.

Ein Benutzer, der sich auf den Inhalt der Kölner Manuskriptbibliothek beziehen will, hat einen Mechanismus zur Verfügung, der "Zitierweisen" für die folgenden Einheiten verlässlich und nachhaltig unterstützt:

1. Digitale Objekte, die die konventionelle bibliographische Bezugseinheit innerhalb der betroffenen Disziplinen wiedergeben, in unserem Fall also einen mittelalterlichen Codex.
2. Digitale Objekte, die eben diese größeren Objekte in einer feineren "Granularität" wiedergeben, die die herkömmliche Zitierweise unterstützt, in unserem Fall individuelle Seiten.

Anmerkung: Wir beziehen uns bewusst zunächst auf "Bezugseinheiten" und "Objektgranularitäten" anstelle von "Codices" und "Seiten". Nicht, um mutwillig eine zusätzliche Ebene von Komplexität einzuführen, sondern um die Verallgemeinerung derartiger Addressierungsschemata ("Zitierweisen") für andere Bereiche des Kulturerbes vorzubereiten. Letzten Endes können Codices als besonders einfacher Fall ge-

sehen werden, wo genau eine Ebene der Unterteilung existiert und die Teilobjekte linear geordnet sind, während beispielsweise Museumsobjekte eine Reihe von Detailebenen aufweisen könnten, die auch durch die Digitalisierung wiedergespiegelt werden muss und bei denen intuitive Systeme für die Benennung der Einheiten in verschiedenen Granularitäten wesentlich komplexer sind. Das hier diskutierte *Grundproblem* liegt jedoch auch dort vor.

Die beiden oben angesprochenen "Zitierweisen" sind aus zwei Gründen notwendig:

1. Aus der Sicht des Benutzers ist es wichtig, ein Zitat auf ein digital gespeichertes Manuskript direkt in ein Werk einzuschließen, das dieses Manuskript zitiert. Dies wird in der Zukunft wesentlich wichtiger werden als heute, wenn Forschungsergebnisse selbst in verstärktem Maße in digitalen Medien publiziert werden. In diesen Fällen wäre es fast schon grotesk, würde der Leser durch eine Fußnote auf die Suchmaschine einer digitalen Bibliothek verwiesen und nicht auf das digitale Objekt selbst, dessen Adresse dem Verfasser zum Zeitpunkt, als er die Fußnote schrieb, ja offenbar bekannt war.
2. Aus der Sicht der bewahrenden Institutionen gibt es einen offensichtlichen Trend hin zu *virtuellen* Bibliotheken, Archiven und Museen. Der einleuchtendste Weg, dieses Ziel zu realisieren, besteht in Zugriffsplattformen, die vor dem Benutzer die Tatsache verstecken, dass die individuellen Objekte unter völlig verschiedenen technischen und administrativen Regelungen gespeichert werden. Und das ist am einfachsten zu realisieren, wenn eine Zugriffsmaschine individuelle digitale Objekte einer Institution direkt ansprechen kann, anstatt sich erst in einen Verhandlungsprozess mit der Software der Institution einlassen zu müssen, die das Objekt verwaltet.

Zumindest in all jenen Bereichen des kulturellen Erbes, in denen die Objekte Einzelstücke sind - zum Unterschied im Wesentlichen von gedruckten Büchern - schie-  
ne es uns extrem unpraktisch, um nicht zu sagen, wirklichkeitsfremd, uns auf die Errichtung einer internationalen Autorität zu verlassen, die einen neuen Satz eindeutiger Identifikatoren für alle weltweit existierenden Objekte des kulturellen Erbes vergibt. Dies würde schließlich letzten Endes darauf hinauslaufen, dass sich alle existierenden Manuskriptbibliotheken, Archive und Museen auf ein einheitliches Signaturesystem einigen. Was wir nicht nur für wirklichkeitsfremd, sondern für unmittelbar schädlich ansehen, da die Verzeichnungssysteme innerhalb unterschiedlicher Sammlungen des kulturellen Erbes, die sich historisch herausgebildet haben,

häufig einen ganz bestimmten intellektuellen Blickwinkel auf dieses Material wiedergeben, der durchaus auch intrinsischen Wert haben kann.

Wir sehen daher eine Lösung vor, die das allgemeine Problem in drei Teilprobleme unterteilt:

1. Ein persistentes Adressierungsschema für *Sammlungen*. Dieses muss notwendigerweise innerhalb der jeweiligen nationalen oder regionalen Verbände organisiert werden und ist mehr eine wissenschaftsorganisatorische, als eine informatikbezogene Frage. Hier können wir daher nur Vorschläge zur Berücksichtigung durch die entsprechenden internationalen Verbände machen.
2. Ein persistentes Adressierungsschema für *digitale Objekte* innerhalb der einzelnen Sammlungen, dessen organisatorische Details unter Kontrolle der jeweiligen Institution stehen, das jedoch, unabhängig von der jeweils verwendeten Software, eine gemeinsame Funktionalität und Interoperabilität der unterschiedlichen Sammlungen garantiert.
3. Ein Abbildungsschema, das es erlaubt eine Untereinheit eines digitalen Objektes durch ein fachspezifisches Zitierschema anzusprechen, das dann in die tatsächlichen Namen der angesprochenen digitalen Komponenten - wie z.B. Seitenabbildungen - übersetzt wird. Solch ein Abbildungsschema wird durch die einzelnen Sammlungen verwaltet und sollte sogar vorgesehen werden, wenn die primären Namen der digitalen Objekte - meist die Dateinamen - die traditionellen Bezeichnungen direkt wiederspiegeln. Die Reihenfolge, in der auf digitale Objekte durch Zugriffsoperationen wie "nächste Seite" zugegriffen wird, ist eine Frage der Interpretation. Um mittelfristig Operationen wie "virtuelles Umbinden" zu unterstützen, schlagen wir nachdrücklich vor, zwischen dieser Ebene und der vorhin erwähnten möglichst detailliert zu unterscheiden.

Eine Implementation der vorangestellten Überlegungen verlangt daher Adressen digitaler Objekte, die formal folgendermaßen aussehen:

<Sammlungsadresse> <Objektadresse> <Objekt-"Granule"-Adresse>

wobei eine <Objekt-"Granule"-Adresse> entweder eine <direkte-"Granule"-Adresse> oder eine <abgebildete-"Granule"-Adresse> sein kann.

Innerhalb von CEEC bieten wir ein arbeitsfähiges Modell für <Objektadressen> und

<Objekt-"granule"-Adressen> an. Für die Diskussion / Definition des Konzepts einer <Sammlungsadresse> sind wir für Anregungen und Unterstützung durch die entsprechenden Verbände dankbar. Alle drei Konzepte sind zum Zeitpunkt dieser Publikation voll funktionsfähig.

## **Diskussion der einzelnen Zugriffsmodelle**

Konkret kann ein vollständiger Kodex der Kölner Dom- und Diözesanbibliothek derzeit durch eine WWW Adresse wie die folgende angesprochen werden:

<http://www.ceec.uni-koeln.de/ceec-cgi/kleioc/0010KICEEC/exec/katk/%22kn28-0083ii%22>

Wenn wir die "%22" (d.h., die WWW-Verpackung der Anführungszeichen) ignorieren, bedeutet dies, dass unsere vorangehenden Definitionen durch folgende Adressierungsbestandteile abgedeckt werden:

<Sammlungsadresse> ::=

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

<Objektadresse> ::=

[ceec-cgi/kleioc/0010KICEEC/exec/katk/%22kn28-0083ii%22](http://www.ceec.uni-koeln.de/ceec-cgi/kleioc/0010KICEEC/exec/katk/%22kn28-0083ii%22)

Um eine individuelle Seite eines Kölner Codex anzusprechen, kann eine WWW Adresse wie die folgende benutzt werden:

[http://www.ceec.uni-koeln.de/ceec-cgi/kleioc/0010KICEEC/exec/pagedmed/%22kn28-0083ii\\_164.jpg%22](http://www.ceec.uni-koeln.de/ceec-cgi/kleioc/0010KICEEC/exec/pagedmed/%22kn28-0083ii_164.jpg%22)

Hier sind folgende Definitionen anwendbar:

<Sammlungsadresse> ::=

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

<Objektadresse> ::=

[ceec-cgi/kleioc/0010KICEEC/exec/pagedmed/](http://www.ceec.uni-koeln.de/ceec-cgi/kleioc/0010KICEEC/exec/pagedmed/)

<Objekt-"Granule"-adresse> ::=

[%22kn28-0083ii\\_164.jpg%22](http://www.ceec.uni-koeln.de/ceec-cgi/kleioc/0010KICEEC/exec/pagedmed/%22kn28-0083ii_164.jpg%22)

### <Sammlungsadresse>

Im Beispiel ist <http://www.ceec.uni-koeln.de> offensichtlich eine URL.

Auf dieser Ebene wären wir für den Dialog mit existierenden Bibliothekseinrichtungen dankbar. Ganz offensichtlich sollte diese URL durch eine ersetzt werden, die besser geeignet ist, langfristig als persistente Identifikation für einzelne Sammlungen zu dienen.

Unserer Ansicht nach wäre dies am besten dadurch zu erreichen, wenn URLs verwendbar würden, die bereits existierende Schemata für die in Frage stehenden Institutionen verwenden. In Deutschland würde dies beispielsweise durch eine URL erreicht, die eine Referenz auf "Kn28" enthält, eine innerhalb der Bundesrepublik eindeutige Bibliothekssigle für die Diözesan- und Dombibliothek Köln.

Weniger förmlich als manche andere dieser Vorschläge: Innerhalb des Internet ist die Frage der verfügbaren Top Level Domains derzeit offen. Mit der Domain \*.museum hat zumindest eine Art von Einrichtung des kulturellen Erbes Top Level Status erreicht.

Wenn wir berücksichtigen, dass auch heute noch die Bibliotheken bei weitem die größten Informationsmengen verwalten, also eigentlich *die* langfristigen Knoten der Informationsnetze sein sollten: Gibt es Bestrebungen die zu weltweit eindeutigen und relativ leicht dauerhaft aufrecht erhaltbaren Bibliotheks-URLs wie "www.kn28.de.lib" führen könnten? Wenn nicht, warum nicht? Gelänge es den Bibliotheksverbänden, auf diese Weise selbst die Kontrolle über die zu vergebenden primären Identifikatoren der Bibliotheken zu erlangen, würde ein nicht unwichtiger Teilkomplex der Frage langfristig dauerhafter Identifikatoren für digitale Ressourcen ganz entscheidend einfacher zu lösen.

### <Objektadresse>

Sobald die Frage der Dauerhaftigkeit der grundlegenden Sammlungsadressen gelöst ist, betrachten wir eine robuste technische Lösung für die weiteren Ebenen als ziemlich einfach.

Wir haben folgendes Schema implementiert::

<Objektadresse>::= <Interface> <Zugriffsmodus> <Objekt-Id>

wobei folgende Überlegungen zu Grunde liegen.

## <Interface>

Das <Interface> einer CEEC <Objektadresse> ist eine Serie von einem oder mehreren Identifikatoren, die durch Schrägstriche getrennt sind. Sie repräsentieren ein Softwaresystem, das zu einem bestimmten Zeitpunkt den Zugriff auf das Material erlaubt. In unserem Beispiel

:

kleioc/0010KICEEC

Anmerkungen:

Die Bezugnahme auf ein spezifisches Interface mag als direkt dem Wunsch nach Dauerhaftigkeit einer URL entgegengesetzt erscheinen. Sie wurde jedoch auf Grund folgender Überlegungen eingeschlossen.

- a) Das einzige, worüber wir über die weitere Entwicklung netzbasierten Informationszuganges in Zukunft sicher sein werden ist, dass er erheblich über den jetzigen Zustand hinausgehen und zusätzliche Möglichkeiten eröffnen wird. Es ist daher äußerst wahrscheinlich, dass digitale Ressourcen in Zukunft Zugriffsmöglichkeiten bieten werden, die es u.U. schwierig machen, genau jene visuellen Effekte sichtbar zu machen, die ein Text, der sich heute auf ein digitales Objekt bezieht, beschreibt.
- b) Andererseits ist es extrem unwahrscheinlich, dass bewahrende Einrichtungen grundlegende Änderungen an ihrer Softwareplattform in häufigen Abständen, häufiger als etwas alle zehn Jahre, durchführen werden.

Wir nehmen daher an, dass eine bibliothekarische Einrichtung, die eine Softwarelösung 'x' in Zukunft durch eine Softwarelösung 'y' ersetzt, eine Möglichkeit haben sollte, durch Unterstützung eines neuen Interfacenamens, bei ansonsten unveränderter Objektadresse, zusätzliche Möglichkeiten zu unterstützen, aber gleichzeitig auch die Möglichkeit haben soll, durch ein entsprechendes Script mit dieser Software unter dem alten Interfacenamen einen Zugriff zu ermöglichen, der den bisherigen Zustand möglichst exakt abbildet.

Da solche Änderungen nicht besonders oft auftreten werden, erscheint es uns als zumutbar, solche Abbildungen auf frühere Zustände von Zeit zu Zeit zu implementieren.

### **<Zugriffsmodus>**

Wir nehmen an, dass nahezu alle vorstellbaren Systeme für die Verwaltung digitaler Objekte den Zugriff auf sie in unterschiedlichen Qualitäten und Auflösungen, bzw. mit unterschiedlichen Zugangs- oder Kostenregelungen unterstützen werden. Der <Zugriffsmodus> einer CEEC <Objektadresse> erlaubt es, bei gleichbleibender restlicher <Objektadresse> gezielt eine bestimmte dieser Qualitäten anzusprechen. Formal gesehen ist er, wie das <Interface> eine durch Schrägstriche getrennte Serie von Identifikatoren. (In unserem Beispiel: "exec/katk").

Anmerkungen:

- a) Wir betrachten es als wichtig, so sauber als möglich zwischen der gewählten Zugriffsart zu einem individuellen Objekt und der Bezugnahme auf dieses Objekt selbst zu unterscheiden.
- b) Die Notation des Zugriffsmechanismus sollte jedoch auch nicht *zu* einfach gestaltet werden. Wir gehen davon aus, dass die internationale Diskussion in absehbarer Zeit zu üblichen Standardqualitäten für digitale Objekte in unterschiedlichen Bereichen führen wird. In diesem Zusammenhang scheint es z.B. wichtig, derartige abstrakte Qualitäten einfach mit davon unabhängigen Zugangsformen kombinieren zu können, die beispielsweise eine bestimmte Gebührenstufe nach der Gebührenordnung der jeweiligen Bibliothek zu Grunde legen.

### **<Objekt-Id>**

Innerhalb des CEEC Mechanismus ist eine <Objekt-Id> eine Zeichenkette, die eine direkte Bezugnahme auf ein digitales Objekt erlaubt, für die ein funktional kompletter Satz von Metadaten existiert, so dass dieses Objekt unabhängig vom Rest der Sammlung angesprochen und ggf. übermittelt werden kann. Für Zwecke besserer Interoperabilität zwischen <Objekt-Ids> unterschiedlicher Sammlungen empfehlen wir nachdrücklich, eine abgekürzte Form der <Sammlungsadresse> in die <Objekt-Id> zu übernehmen. Dadurch sollte es im Prinzip möglich sein, aus der <Objekt-Id> allein eine vollständige URL zum Zugriff auf dieses Objekt abzuleiten.

### **<Objekt-"Granule"-Adresse>**

Innerhalb des CEEC Mechanismus ist eine <Objekt-"Granule"-Adresse> eine Zei-

chenkette, die eine direkte Bezugnahme auf die kleinste unabhängig adressierbare Einheit digitalisierter Information innerhalb des digitalisierten Objektes anspricht. In aller Regel wird dies die Datei sein, die die Daten einer gescannten Seite enthält. Für Zwecke besserer Interoperabilität zwischen <Objekt-"Granule"-Adressen> unterschiedlicher Sammlungen empfehlen wir nachdrücklich, dass die vollständige <Objekt-Id> in der <Objekt-"Granule"-Adresse> wiederholt wird. Dadurch sollte es im Prinzip möglich sein, aus der <Objekt-"Granule"-Adresse> allein eine vollständige URL zum Zugriff auf diese Seite abzuleiten.

### <direkte-"Granule"-Adresse>

Sie besteht aus einer Zeichenkette, die verwendet werden kann, um direkt auf die digitalisierte Information auf einem spezifischen Server zuzugreifen. Das bedeutet, dass sie für den Zugriff in Komponenten zerlegt werden kann, die unterschiedliche Ebene einer Speicherhierarchie ansprechen und / oder logische Namen auf physikalische Geräteadressen abbilden. Sie erlaubt jedoch *keine* konzeptuelle Interpretation ihrer Bestandteile. Die Institution, die die digitale Sammlung realisiert, garantiert, dass die <direkte-"Granule"-Adresse> einer digitalisierten Seite, bzw. eines anderen atomischen digitalen Objektes, sich während seiner Existenz nie ändert.

In unserem Beispiel: kn28-0083ii\_164.jpg

### <abgebildete-"Granule"-Adresse>

Sie besteht aus einer Zeichenkette, die durch ein Trennzeichen in zwei Teile zerlegt wird - wir verwenden innerhalb der CEEC Implementation dafür den senkrechten Strich "|". Der erste der beiden Teile identifiziert einen spezifischen Mechanismus durch den der zweite Teil auf eine <direkte-"Granule"-Adresse> abgebildet werden kann und zwar nach Regeln, die sich während der Lebenszeit des digitalen Objektes ändern, die aber auch als überholt gelöscht werden können.

Wenn eine <abgebildete-"Granule"-Adresse> mit einer vertikalen Linie *beginnt*, bedeutet dies andererseits, dass die Abbildung durch einen Standardmechanismus geschieht, der während der gesamten Lebenszeit des digitalen Objektes unterstützt wird und eine "*kanonische Referenz*" genannt wird.

In unserem Beispiel:

|kn28-0083ii\_82r wird auf jene Datei abgebildet, die Seite 82 *recto* des Manuskripts nach der üblichen Zählweise der Literatur repräsentiert.

Müller|kn28-0083ii\_einschub4-3r kann beispielsweise abgebildet werden in die Datei, die Seite 3 *recto* des vierten Einschubs in ein hypothetisches Originalmanuskript repräsentiert, nach der vom Forscher "Müller" vorgeschlagenen Zählweise. Diese Interpretation kann, in Abhängigkeit vom weiteren Schicksal dieser These geändert werden oder auch als unauflösbar zurückgewiesen werden, wenn sich herausgestellt hat, dass diese These nicht haltbar ist. (Ein Zugriff nach diesem letzten Beispiel ist derzeit noch *nicht* realisiert.)