

Die Praxis der Digitalisierung

von TORSTEN SCHASSAN

Nach den Beschreibungen inhaltlicher und formaler Aspekte in dem Projekt CEEC soll an dieser Stelle nun die Praxis folgen. Dabei wird der komplette Ablauf dargestellt, von der Auswahl und Reihenfolge der Manuskripte zur Digitalisierung über die eigentliche Digitalisierung bis hin zur Weiterverarbeitung und zum Einspielen der Scans ins Internet.

1. Auswahl der Handschriften

Die Auswahl der Handschriften geschah in der ersten Projektphase (die ersten 3-4 Monate) vor allem unter dem Gesichtspunkt, einen Querschnitt aus dem Bestand der DDB in Hinsicht auf Alter, Inhalt, Schriftentwicklung etc. zu erstellen. Es wurde in Zusammenarbeit mit dem damaligen, jetzt pensionierten Direktor der Erzbischöflichen Diözesan- und Dombibliothek (DDB), Herrn Cervelló Margalef, dem Leiter der Handschriftenabteilung, Herrn Lenz und dem Konservator der Bibliothek, Herrn Schäfers, eine Liste der Auswahlkriterien, sowie auch eine konkrete Liste der zu digitalisierenden Handschriften erstellt. Die Auswahlkriterien für diese erste Phase waren im Einzelnen die folgenden:

- **Bibliotheksorganisation**

Bei allen Handschriften galt und gilt, dass Rücksicht auf die Bibliotheksorganisation genommen wird. Dies kann sich auf Ausleihen, Restaurationsarbeiten (Entbindungen etc.) oder etwaige andere Vorbehalte beziehen. Die gesamte Auswahl, besonders die der im weiteren Verlauf des Projektes anstehenden großformatigen Manuskripte, erfolgt in Zusammenarbeit mit dem Konservator der Bibliothek

- **Erhaltungszustand**

Zuerst sollten die vom Erhaltungszustand "unproblematischen" Handschriften digitalisiert werden. Unproblematisch bedeutet in diesem Zusammenhang, dass die Handschriften keine größeren Schäden aufweisen, welche die Handhabung und Digitalisierung erschweren würden. Ggf. sollten beschädigte Handschriften

sowohl aus Übungszwecken bei den Digitalisierungsarbeiten wie auch zur Demonstration möglicher Schäden und zur Entwicklung spezieller Bildbearbeitungswerkzeuge vorgezogen werden.

- **Thematischer Warenkorb**

Es sollte eine beispielhafte Auswahl der Handschriften aus allen Inhaltsbereichen getroffen werden, um den Benutzern frühzeitig einen inhaltlichen Überblick über die Bestände der Bibliothek zu ermöglichen. Zu den wichtigsten Inhaltsbereichen zählen Kanonistik, Patristik, Komputistik, "Literatur", Philosophie und Bücher für die Liturgie.

- **Fachwissenschaftliches Interesse**

Vorrang wurde den Handschriften eingeräumt, bei denen aufgrund fachwissenschaftlichen Interesses eine kooperative Erschließung zu erwarten ist. (Beispiele: theologisch, kodikologisch, paläographisch, historisch, kunsthistorisch, philologisch). Vgl. dazu auch unten die Erläuterungen zu möglichen Kooperationen.

- **Alter**

Ältere Handschriften sollten vor jüngeren Handschriften digitalisiert werden.

- **Bibliothekszugehörigkeit**

Der Bestand der Dombibliothek sollte vor dem der Diözesanbibliothek bearbeitet werden. Damit wird, wie auch mit der Berücksichtigung des Alters, der speziellen Geschichte der Dombibliothek Rechnung getragen, deren Manuskripte zu einem Großteil dem frühen Mittelalter entstammen und bereits im 9. Jahrhundert in der Dombibliothek vorhanden waren. Zahlreiche Manuskripte enthalten den Hinweis auf den ersten Kölner Erzbischof Hildebold (Hildebald).

- **Herkunft**

Stadtkölnische Provenienz, spezielle Schreibschulen und Handschriften von besonders seltene Herkunft sollte vorrangig bearbeitet werden.

- **Illuminationen / Illustrationen**

Illuminierte Handschriften sollten vor nicht-illuminierten Handschriften bearbeitet werden, zunächst allerdings nur beispielhaft, um z.B. mit Goldauflagen Erfahrungen zu sammeln. Spezielle Einstellungen bei den Aufnahmen ließen einen größeren Zeitaufwand pro Aufnahme erwarten. Anspruchsvollere Kodizes wer-

den seit dem Ende der ersten Projektphase (mit besseren Erfahrungswerten) digitalisiert.

- **Format**

Anfangs sollten hauptsächlich "normal"-formatige Handschriften bearbeitet werden. Als normales Format gilt ungefähre DIN A4 Größe. Die obere Grenze liegt bei ca. 40 mal 30 cm, die untere bei ca. 21 mal 15 cm. In der anbrechenden zweiten Projektphase (Monate 6-12) werden immer mehr auch Großformate digitalisiert, um rechtzeitig die auftauchenden Probleme durch das hohe Gewicht und die niedrigere Auflösung berücksichtigen und Vorausplanungen machen zu können.

1.1. Zusammenarbeit

Wenn einzelne Forscher besonderes wissenschaftliches Interesse an einer Handschrift haben und diese noch nicht digitalisiert ist, können sie die Projektmitarbeiter kontaktieren. Es wird dann geprüft, ob das betroffene Manuskript in der vorgesehenen Reihenfolge vorgezogen werden kann. Darüber hinaus ist für die weitere Erschließung ein System von Patenschaften in Planung: Forscher, die mit oder über eine Handschrift der DDB gearbeitet oder Editionen der darin befindlichen Texten erstellt haben, sollen für die Mitarbeit gewonnen werden. Dabei entstehende Texte, die z.B. die Forschung über die Handschrift zusammenfassen oder eine synopsenartige Beschreibung liefern könnten, würden als zitierfähige Literatur in das Datenbanksystem eingebaut werden. Modelle einer weiteren Veröffentlichung sind denkbar.

2. Die Digitalisierung

2.1. Die technische Ausstattung

Die technische Ausstattung in der Bibliothek besteht aus zwei Pentium III-Computern, einer Digitalkamera und einem Kameraarm nebst Leuchtmitteln.

Ein Computer bedient und steuert die Kamera und speichert die Scans übergangsweise (Kamerarechner), der andere ist für die erste Weiterverarbeitung und die Speicherung der Daten auf CD-ROM vorgesehen (Brennerrechner). Die Rechner sind über ein einfaches Netzkabel miteinander verbunden. Auf dem Kamera-

rechner ist Windows-NT installiert, auf dem Brennerrechner Windows 98. Beide Rechner haben zwei Festplatten von je 20 GB Größe, der Brennerrechner zusätzlich eine weitere 40 GB Festplatte. Auf dem Kamerarechner ist die Software Imagic installiert, welche die Kamera steuert.

Die Digitalkamera ist eine ProgRes 3012, früher bekannt als Kontron. Das Lamegon Objektiv hat eine Brennweite von 14mm und eine Lichtstärke von 3,5. Die maximale Auflösung der Kamera beträgt 4491 mal 3480 Punkte.

Da die zu digitalisierenden Handschriften überaus empfindlich sind, insbesondere gegenüber UV-Licht und Wärme bzw. Luftfeuchtigkeit, werden die Digitalisierungsarbeiten mit besonderen Leuchtmitteln durchgeführt. Die verwendeten Lampen sind extrem kalt, d.h. die Erwärmung ist minimal. Sie liegt dauerhaft bei 1-2 Grad Celsius Erwärmung gegenüber der Raumtemperatur. Außerdem ist das Licht der Lampen fast UV-frei. Ein Schaden für die Handschriften von dieser Seite sollte damit ausgeschlossen sein.

Die Kamera und Leuchtmittel sind an einem einfachen Kameraarm befestigt.

Die Digitalisierung erfolgt immer mit der höchsten Auflösung von 4491 mal 3480 Punkten. Je nach Format der Handschrift beträgt die Auflösung der Bilder deshalb zwischen 200 und 600 PPI.¹ Das Speicherformat ist TIFF. Die Dateien sind unkomprimiert jeweils 45 MB groß. Eine Aufnahme dauert ca. 1 Minute, weitere 20 Sekunden das Speichern des Scans. Um die ca. 130.000 Seiten bzw. 65.000 Seiten im Antragszeitraum in der vorgegebenen Zeit digitalisieren zu können, muss im Rahmen der Öffnungszeiten ein durchschnittliches Ergebnis von 17 Bildern pro Stunde erzielt werden.

Angesichts der langen Belichtungszeit für die Aufnahme ist es notwendig, die Handschriften zu fixieren. Dies geschieht im Projekt mit Hilfe einer auf die zu digitalisierenden Handschriftenseite aufgelegten Glasplatte. Je nach Größe der Handschrift hat auch die Glasplatte ein unterschiedliches Format. Diese Glasplatten sind speziell entspiegelt, um eine möglichst hohe Farbreinheit und wenig Reflexe zu gewährleisten. Zudem sind die Platten extrem dünn. Mit nur ca. 2 Millimeter Glasstärke entsteht dadurch keine Belastung für die Handschriften. Die mechanischen

¹ "PPI" ist die Abkürzung für "Points Per Inch". PPI wird meist synonym gebraucht mit DPI (=Dots Per Inch), welches die Angaben für einen Drucker sind. Hat eine Handschrift z.B. das Format 20 mal 15 cm, dann ergibt sich daraus bei der Maximalauflösung von 4491 mal 3480 Punkten ein Auflösungswert von ca. 560 PPI. (20 cm entsprechen 7,874 Inch. 4491 geteilt durch 7,874 = 570,36) Bei einer Handschrift der Größe 40 mal 30 cm liegt dieser Wert "nur noch" bei ca. 285 PPI. Die notwendige Auflösung, um z.B. eine gute Qualität bei Ausdrucken zu erzielen, liegt bei mindestens 300 PPI. Damit würden die Großformate unter diese Marke rutschen.

Belastungen alleine durch das Aufklappen eines Manuskriptes sind im Vergleich sicherlich schädlicher.

Um die Handschriften zu schonen, werden sie möglichst nicht über 110 Grad geöffnet. Die aufgeschlagenen Seiten lehnen gepolstert gegen den Kameraarm. Das Manuskript wird bei Bedarf mit Schaumstoffpolstern gestützt, um z.B. das Gewicht der aufgeschlagenen Seiten aufzufangen oder die zu digitalisierende Seite der Handschrift in waagerechte Position zu legen.

Die Anschaffung einer Buchwippe wird derzeit in Betracht gezogen, doch sind bisher getestete Wippen bei den großformatigen Handschriften aufgrund des dann sehr hohen Gewichtes unbrauchbar.

2.2. Die personelle Ausstattung

Die Digitalisierungsarbeiten werden von einem externen Dienstleister durchgeführt. Die Firma Image Engineering aus Frechen beschäftigt zu diesem Zweck 4 studentische Hilfskräfte, die in den Öffnungszeiten der Bibliothek (42 Stunden pro Woche) die Kamera bedienen und die weiteren notwendigen Arbeiten übernehmen. Zwei der Studenten sind Studenten des Faches Informationsverarbeitung, die anderen beiden studieren an der Fachhochschule Köln Photoingenieurwesen. Die Auswahl und Einarbeitung der Studenten geschah in Zusammenarbeit und in Absprache mit den Projektleitern. Die Studenten werden von den wissenschaftlichen Mitarbeitern regelmäßig, von dem Konservator der Bibliothek stichprobenartig betreut und kontrolliert.

Die Studenten haben zu gewährleisten, dass die Aufnahmen scharf sind. Bei der Ausbildung für diese Tätigkeit wurden entsprechende Referenzen und Vergleichsmöglichkeiten gegeben. Eine Kontrolle der Digitisate findet beim Einspielen der Scans ins Internet durch eine weitere studentische Hilfskraft und die wissenschaftlichen Mitarbeiter statt. Nicht gelungene Aufnahmen werden reklamiert und wiederholt.

Von der personellen wie technischen Ausstattung her ist das Projekt "Codices Electronici Ecclesiae Coloniensis" ein eher kleines Projekt. Ein Ziel des Projektes ist es daher auch, ein Beispiel dafür zu geben, dass mit diesen geringen Mitteln und relativ wenig Aufwand durchaus achtbare Ergebnisse zu erzielen sind.

2.3. Regeln bei der Aufnahme

Im Projekt werden in den ersten beiden Jahren insgesamt 65.000 Seiten digitalisiert. Diese "Massenproduktion" erfordert bestimmte Grundregeln bei der Bearbeitung. Das wichtigste Ziel ist die gleichbleibende Qualität der Aufnahmen. Gleichbleibend bezieht sich sowohl auf den Ausschnitt, der zu sehen ist, als auch auf die Farbgebung und die Schärfe. Ein Kodex wird möglichst mit *einer* Einstellung der Blende und des Lichtes aufgenommen, Veränderungen dieser "Grundeinstellungen" bei einem Kodex sollten nicht vorgenommen werden. Auf den Bildern sollten deshalb die Farb- und Helligkeitsunterschiede des Beschreibstoffes selbst zu erkennen sein. So unterscheiden sich z.B. die Haar- und Fleisch-Seite des Pergaments: die Haar-Seite, also die "Außenseite" des Felles ist dunkler gefärbt als die Fleisch-Seite.

Um einen Vergleichsmaßstab zu haben und evtl. automatische Farbkorrekturen vornehmen zu können, wird eine Farbskala mit digitalisiert.

Bei der Digitalisierung soll die komplette Seite zu sehen sein, d.h. anders als bei vielen Photographien wird der Rand mit abgebildet. Der Untergrund ist ein Raster, an dem die Größe der jeweiligen Handschrift erkennbar ist.

2.4. Regeln zur Benennung der Scan-Dateien

2.4.1. Grundsätzliches

Grundsätzlich ist bei der Benennung der Dateien folgender Regelname zu vergeben: **Kn28-xxxx_yyy_zo.tif**

Kn28 ist das Bibliothekskennzeichen der Erzbischöflichen Diözesan- und Dombibliothek im Deutschen Bibliotheksverbund. Damit ist eine eindeutige Identifizierung der Bibliothek möglich. Sollten später einmal mehrere Bibliotheken als virtuelle Bibliotheken existieren, lässt sich dennoch die Zugehörigkeit der Bücher zu einer der Bibliotheken immer erkennen.

xxxx steht für die Signatur einer Handschrift in der Bibliothek. Handschriften der Dombibliothek werden mit einer dreistelligen Zahl und einer vorgestellten Null gekennzeichnet, Handschriften der Diözesanbibliothek mit einer dreistelligen Zahl und einer vorgestellten Eins. Dies entspricht der bibliotheksinternen Praxis.

yyy steht für die Seitennummern. Die Vorderseite des ersten nummerierten Folio der Handschrift ist dementsprechend Kn28-xxxx_001_zo.tif. Im Folgenden

werden die Bilder fortlaufend nummeriert. Eine Benennung der Scans entsprechend den mittelalterlichen Regeln der Foliierung und der daraus resultierenden Bezeichnung mit recto- und verso-Seiten erfolgt nicht. Folio 1r wird also Kn28-xxxx_001_zo.tif, Folio 1v Kn28-xxxx_002_zo.tif, Folio 2r Kn28-xxxx_003_zo.tif, etc

zo steht für die Ausrichtung der gescannten Seite. Zur besseren Ausleuchtung der Handschriftenseite werden die Handschriften quer zu den Leuchtmitteln aufgeschlagen. Die Kamera liefert ein querformatiges Bild und dementsprechend erscheinen die Handschriften nicht hochkant auf dem Ursprungsbild. Je nachdem ob die obere Kante der Handschrift links oder rechts auf dem Scan zu sehen ist, kann Z die Werte "r" und "l" annehmen. Ein Scan, auf dem die Handschrift im Uhrzeigersinn um 90° gedreht, also mit der oberen Kante rechts im Bild erscheint, würde mit "ro" gekennzeichnet.

2.4.2. Ausnahmen:

2.4.2.1. Der Einband:

Aufnahmen des Einbandes werden nach der Regel benannt, dass der Bildnummer der Aufnahme der Buchstabe "e" vorangestellt wird, stellvertretend für "Einband".

- Die Aufnahme Kn28-xxxx_e001_zo.tif ist ein Bild der Vorderseite des Einbandes.
- Die Aufnahme Kn28-xxxx_e002_zo.tif ist ein Bild des Buchrückens.
- Die Aufnahme Kn28-xxxx_e003_zo.tif ist ein Bild der Rückseite des Einbandes.
- Die Aufnahme Kn28-xxxx_e004_zo.tif ist ein Bild der Schließenseite des Einbandes.
- Die Aufnahme Kn28-xxxx_e005_zo.tif ist ein Bild der vorderen Innenseite des Einbandes.
- Die Aufnahme Kn28-xxxx_e006_zo.tif ist ein Bild der hinteren Innenseite des Einbandes.

• Nicht nummerierte Seiten:

Nicht nummerierte Blätter einer Handschrift (vorgeheftet, eingelegt etc.) werden folgendermaßen benannt:

Vorgeheftete Blätter erhalten fortlaufende Nummern entsprechend dem Schema:

- Kn28-xxxx_000a_zo.tif für die Vorderseite des ersten nicht nummerierten Blattes;
- Kn28-xxxx_000b_zo.tif für die Rückseite des ersten nicht nummerierten Blattes;
- Kn28-xxxx_000c_zo.tif für die Vorderseite des zweiten nicht nummerierten Blattes, etc.

Nachgeheftete Blätter erhalten fortlaufende Nummern entsprechend den Schema:

- Kn28-xxxx_999a_zo.tif für die Vorderseite des ersten nicht nummerierten Blattes;
- Kn28-xxxx_999b_zo.tif für die Rückseite des ersten nicht nummerierten Blattes;
- Kn28-xxxx_999c_zo.tif für die Vorderseite des zweiten nicht nummerierten Blattes, etc.

Eingelegte Blätter erhalten fortlaufende Nummern entsprechend den Schema:

- Kn28-xxxx_yyya_zo.tif für die Vorderseite der ersten nicht nummerierten Seite;
- Kn28-xxxx_yyyb_zo.tif für die Rückseite der ersten nicht nummerierten Seite;
- Kn28-xxxx_yyyc_zo.tif für die Vorderseite der zweiten nicht nummerierten Seite.

yyy ist die Nummer der Rückseite des Blattes vor dem eingelegten Blatt.

2.4.3. Ausschnittvergrößerungen

Ausschnitte erhalten fortlaufende Nummern entsprechend dem Schema:

- Kn28-xxxx_ayyy_zo.tif; yyy ist die Seitenzahl, von welcher der Ausschnitt aufgenommen wurde.

• Die Weiterverarbeitung der Scan-Dateien

Wie bereits erwähnt bedient der Kamerarechner die Kamera und speichert die Bilder kurzfristig, auf dem Brennerrechner wird die erste Weiterverarbeitung und Langfristsicherung durchgeführt. Die Scans werden über die Netzwerkverbindung vom Kamerarechner auf den Brennerrechner übertragen und durchlaufen dort einen Batchjob, der von ausgeführt wird.² Dabei werden die Bilder komprimiert³ und gedreht. Wie oben bei den Benennungsregel bereits angedeutet, werden die Hand-

² Ein Batch (engl. Stapel, to batch = stapelweise verarbeiten) ist eine Abfolge festgelegter Bearbeitungsschritte, die mit jedem einzelmem Scan durchgeführt werden. Κλειω (oder: kleio) ist ein an Vorstellungen semantischer Netzwerke orientiertes, nicht-relationales Datenbankverwaltungssystem. Es wurde vom Projektleiter Prof. Dr. Manfred Thaller entwickelt.

³ Das Komprimierungsverfahren heißt LZW-Komprimierung. Es ist nach den Erfindern Lempel, Ziv und Welch benannt und ermöglicht die informationsverlustfreie Komprimierung. Informationsverlustfrei bedeutet, dass nur die einzelnen Bildpunkten zusammengefasst werden, die wirklich identische Informationen wie Farb- oder Helligkeitswerte tragen.

schriften nicht hochkant aufgenommen und müssen um 90 bzw. 270 Grad gedreht werden.

Der Batchjob wird in der Regel bei Ende der Öffnungszeiten der Bibliothek in Gang gesetzt und kann am Abend abgearbeitet werden. Ein Ausschaltmechanismus sorgt dafür, dass die Computer nachts abgeschaltet sind.

Die endgültigen Scans werden in zweifacher Ausfertigung auf CD-ROM gebrannt. Eine CD-ROM verbleibt in der Bibliothek, das zweite Exemplar wird in der Universität weiterverarbeitet.

3. Endverarbeitung der Scans

In der Universität werden von jeder Datei vier Versionen unterschiedlicher Größe angefertigt, im jpg-Format gespeichert und auf dem Server des Projektes abgelegt. Der dafür notwendige Plattenplatz wird von dem Regionalen Rechenzentrum der Universität zu Köln bereitgestellt.

Im Web stehen diese vier (in Ausnahmefällen 5) Varianten einer Handschriftenseite zur Verfügung. Diese werden durch die folgenden Icons symbolisiert:

-  Das Icon "Auge" steht für die kleinste Variante, die es auch auf einem Bildschirm mit der Auflösung 800*600 Pixel möglich machen soll, die gesamte Handschriftenseite auf einen Blick zu sehen. Es erschien uns wichtig, die Möglichkeit eines schnellen Überblicks zu geben. Von dort aus kann man in mehreren Stufen in das Bild "hineinzoomen".
-  Das Icon "Brille" symbolisiert ein Bild mit einer Breite von 1000 Pixel, so dass die volle Breite einer Zeile Schrift auf einem Monitor mit der Auflösung 1024*768 Pixel dargestellt werden kann.
-  Die dritte Auflösung mit einer Bildgröße von 1600 Pixel Breite, dargestellt durch die "Lupe", ist eine technisch veränderte Datei. Der Kontrast wird erhöht, um so die Lesbarkeit des Textes zu verbessern.
-  Sollte eine Stelle im Text immer noch nicht zu entziffern sein, kann noch die höchste Auflösung abgerufen werden, die durch das Symbol "Mikroskop" vertreten wird. Diese Datei liegt in originaler Auflösung von 4491 mal 3480 Pixel vor, allerdings auch hier im jpg-Format.
-  Zu einigen Stellen einer Handschriftenseite kann es zusätzliche Detailaufnahmen geben, die durch diesen Button aufgerufen werden können.

Die Bilddateien sind jeweils mit einem Wasserzeichen versehen. Die Originaldateien können nach der Unterzeichnung eines Nutzungsvertrages bei der Bibliothek auf CD-ROM angefordert werden.

Um die neu digitalisierten Handschriften der Datenbank bekannt und zugänglich zu machen, wird über kleine Programme geprüft, welche Seiten aus welchen Handschriften digitalisiert vorliegen. Die XML-Daten werden aktualisiert und die entsprechenden, notwendigen Links generiert.

Für den Arbeitsablauf vom Brennen der CD-ROMs bis zum Einspielen ins Internet und der Qualitätskontrolle gibt es einen "Laufzettel".

Der Zugriff auf Bilder und Katalog- bzw. Metadaten geschieht mit Hilfe von κλειϖ. Über eine CGI-Schnittstelle werden die HTML-Seiten, die Bilder und Metadaten enthalten können, automatisch von der Datenbank generiert.