
Digitale Bibliotheken in den USA

Ein Reisebericht

Alexander Egger, Elisabeth Schloffer

Im Februar dieses Jahres hatte ich im Anschluss an eine Konferenz in New York dank unserer internationalen Kontakte aus dem METAe-Projekt die Möglichkeit, einige Institutionen in den USA, in denen Bibliotheksbestände digitalisiert werden, zu besuchen. Ziel war es, für unser METAe-Projekt die tägliche Praxis in der Digitalisierung besser kennen zu lernen.

Reiseroute

Auf meiner zehn Tage langen Reise beschränkte ich mich auf die vier „Must Sees“ der digitalen Bibliotheken. Meine Reise begann in New York in der *New York Public Library*, welche vor allem für ihre Digitalisierungsabteilung bekannt ist. Weiter ging es dann nach Detroit in die *University of Michigan*, deren Stärke in der Erstellung und Verwaltung von elektronischen Texten liegt. Meine vorletzte Station, die *Cornell University in Ithaca* in der Nähe von Pittsburgh, entwickelt Software für digitale Bibliotheken. *Harvard University, Boston*, - letzte Station meiner Reise - ist bekannt für ihr hervorragend eingerichtetes Digitalisierungslabor.

Ziele der Reise

Wie schon erwähnt, ging es uns vor allem darum, die tägliche Praxis der Digitalisierung kennen zu lernen. Die Schwerpunkte setzte ich dabei auf folgende Punkte:

- Erstellung von digitalen Sammlungen
- Verwaltung von digitalen Sammlungen
- Eingesetzte Technologien
- Organisationsformen
- Zukünftige Entwicklungen

New York Public Library

Die New York Public Library besteht aus 4 Forschungszentren in Manhattan und 85 Zweigstellen in der Bronx, in Manhattan und Staten Island. Die wissenschaftlichen Bibliotheken halten etwa 42 Millionen Objekte, davon etwa 14 Millionen Bücher. Die 85 Zweigstellen halten weitere 11 Millionen Objekte, 4 Millionen davon sind Bücher. Jährlich benutzen mehr als 2 Millionen Besucher die Forschungsbibliotheken und etwa 13 Millionen die Zweigstellen. Die Internet-Seite der New York Public Library verzeichnet ca. 3,6 Millionen Besucher jährlich. Die New York Public Library ist damit die größte öffentliche Bibliothek der Vereinigten Staaten und eine der größten wissenschaftlichen Bibliotheken der Welt.

Bei meinem Besuch waren für mich vor allem zwei Abteilungen interessant: die „Digital Imaging Group“, die sich mit der Digitalisierung von Objekten beschäftigt, und die „Digital Library Collection“, welche die digitalen Sammlungen der Bibliothek aufbaut und betreut.

Digital Imaging Group

Die „Digital Imaging Group“ erzeugt digitale Faksimiles höchster Qualität von Buchseiten, Bildern, Landkarten etc. Für die Digitalisierung werden Flachbettscanner beziehungsweise eine digitale Kamera eingesetzt. Die digitale Kamera wird für Objekte eingesetzt, die aufgrund ihrer Größe nicht mit dem Flachbettscanner digitalisiert werden können.

Als minimaler Standard gilt eine Auflösung von 300 dpi und eine Farbtiefe von 24 Bit. Besonderer Wert wird auf die Farbtreue gelegt. Daher ist die gesamte Produktionsstraße, vom Scanner, über die Monitore bis zum Drucker kalibriert. Als Farbraum wird der Kodak proRGB-Farbraum verwendet. Mit jedem Foto wird auch ein Kodak Standard-Farbstreifen mit aufgenommen.

Der Farbstreifen wird alle vier Wochen ausgetauscht, um einem Ausbleichen des Streifens durch die Scheinwerfer vorzubeugen. Beim Scannen wird für jede produzierte CD-ROM ein Farbstreifen gescannt und auf der CD-ROM mitgespeichert. Derzeit werden alle Bilder auf CD-ROM gespeichert, für die Zukunft ist ein 18 Terrabyte-Speicher geplant. Jede Aufnahme wird einer genauen Qualitätskontrolle unterzogen bevor sie gespeichert wird.

Für die Darstellung im Internet werden die Bilder zusätzlich in ein proprietäres Wavelet-Format umgewandelt. Dieses Format garantiert zusammen mit einer eigenen Viewing-Software (Luna Insight), dass die Benutzer zwar beliebige Ausschnitte in beliebiger Qualität betrachten können, nie aber das gesamte Bild in der höchsten Qualität. Auf diese Weise werden die Rechte auf das digitale Bild geschützt.

Das derzeit größte Projekt „Digital Treasures“ wird in den nächsten Jahren über 600.000 digitale Faksimiles erzeugen.

Digital Library Collection

Die „Digital Library Collection“ übernimmt die Verwaltung von digitalen Sammlungen der New York Public Library. Den Hauptteil der Sammlung bilden sogenannte „Finding Aids“, elektronische Texte, die diverse Sammlungen der Bibliothek beschreiben. Zur Zeit gibt es 300 solcher „Finding Aids“. Diese Texte werden durch 150 elektronische Bücher und die Faksimiles der „Digital Imaging Group“ ergänzt.

Alle elektronischen Texte der Bibliothek werden im TEI-Format gespeichert. Die Software DynaWeb übernimmt die Darstellung der Texte über HTML-Dateien im Internet.

University of Michigan

Die Universität von Michigan wurde 1817 gegründet. Heute besuchen am Campus von Ann Arbor etwa 37.000 Studenten mehr als 5600 verschiedene Kurse.

Die Universitätsbibliothek besteht aus 25 - über den Campus verstreute - Bibliotheken. Ein einheitliches elektronisches Katalogsystem verbindet die einzelnen Bibliotheken.

Die Universitätsbibliothek bietet ein großes Spektrum an Sammlungen verschiedener Medien. Neben Büchern und Zeitschriften werden auch Musik, Film, Video und eine große Anzahl an elektronischen Ressourcen angeboten.

Elektronische Zeitschriften und Volltextversionen von literarischen und historischen Werken werden mit 30.000 elektronischen Zitaten zu Zeitschriften er-

gänzt. Ein Netzwerk mit 10 anderen amerikanischen Universitäten ermöglicht den Zugriff auf 60 Millionen digitale Objekte.

Zwei Institute der Universitätsbibliothek beschäftigen sich mit der Verwaltung und der Erstellung von digitalen Ressourcen - das „Digital Library Production Service“ und das „Digital Library Project“.

Digital Library Production Service

Das „Digital Library Production Service“ wurde 1996 gegründet, um eine Infrastruktur für die digitalen Sammlungen der Universität zu schaffen. Beide Aspekte - Digitalisierung von Objekten und der Zugang zu diesen Objekten - werden von diesem Service abgedeckt. Die Grundidee ist es, mit einem zentralen Service für alle Bibliotheken einen Wildwuchs an verschiedenen Systemen innerhalb des Campus zu vermeiden. Durch das zentrale Service bleibt die Anzahl der eingesetzten Dateiformate und Systeme überschaubar, die digitalen Sammlungen sind besser wartbar und ein Verlust der Daten kann besser verhindert werden.

Trotz des zentralisierten Ansatzes hat jede Bibliothek die Freiheit ihre digitalen Sammlungen sowohl inhaltlich als auch bezüglich der Präsentation frei zu gestalten.

Die digitalen Sammlungen bestehen hauptsächlich aus SGML-Texten, Faksimiles von Journalen, Bildern von Museumsobjekten und numerischen Datensammlungen.

Technisch gesehen übernimmt das „Digital Library Production Service“ folgende Aufgaben:

- Speicherung der Daten
- Indizierung der Daten
- Management für Digital Library Projekte
- Entwicklung und Wartung von Software

Das „DLPS“ verwaltet zur Zeit 1,6 Millionen Faksimileseiten, 5,2 Millionen Volltextseiten und 2 Millionen bibliographische Datensätze.

Untereinheiten des „Digitale Library Production Service“

Das „Digitale Library Production Service“ teilt sich in zwei Einheiten, die Digitalisierungseinheit und die „Information Retrieval/Architecture“-Einheit.

Die Digitalisierungseinheit erstellt mit einem OCR-Cluster Volltext im großen Stil (mehrere Millionen Seiten pro Jahr). Die Digitalisierung der Buchseiten und die XML-Kodierung des Volltexts wird von Vertragsfirmen erledigt. Eine

kleine Untereinheit der Digitalisierungseinheit beschäftigt sich mit der Erstellung von hochgenauen Faksimiles von Museumsobjekten. Dazu wird ein Bild mit einer analogen Kamera angefertigt und der Film dann mit einem Film-scanner digitalisiert.

Die „Information Retrieval/Architecture“-Einheit erstellt digitale Bibliotheken und übernimmt deren Wartung. Die Daten für diese Bibliotheken sind SGML- oder XML-Volltextdateien, digitale Faksimiles oder numerische Daten.

Digital Library Project

Mit dem „Digital Library Project“ versucht die Universität Michigan einen neuen Weg für digitale Bibliotheken zu gehen.

Bibliotheken bieten historisch gesehen ein organisiertes, zentralisiertes Service an. Das „World Wide Web“ im Gegensatz dazu ist ein dezentrales, unorganisiertes Service.

Digitale Bibliotheken, mit traditionellen Bibliotheken als Hintergrund, bieten ein zentralisiertes Service und verlieren dadurch die Vorteile des WWW, wie Flexibilität und Dezentralisierung.

Das „Digital Library Projekt“ versucht den Benutzern das Gefühl einer traditionellen Bibliothek zu vermitteln. Das Service soll aber auf einer dezentralen, flexiblen Infrastruktur aufbauen. Den Kern des neuen Systems werden Softwareagenten bilden, die für jeden Benutzer Profile anlegen und ihn aufgrund dieser Profile bei seiner Arbeit unterstützen.

Das Projekt befindet sich allerdings erst in seiner Anfangsphase.

Cornell University

Die Cornell Universität wurde 1865 gegründet. Heute besuchen etwa 20.000 Studenten diese Universität.

Die Universitätsbibliothek besteht - wie an der Universität von Michigan - aus verschiedenen einzelnen Bibliotheken. Das Bibliothekssystem von Cornell umfasst 19 Bibliotheken; insgesamt werden 7 Millionen Bücher verwaltet.

In Cornell beschäftigt sich eine Vielzahl von Institutionen mit dem Aufbau und der Verwaltung von digitalen Sammlungen. Die „Digital Library Research Group“ und das „Cornell Institute for Digital Collections“ sollen hier genauer beschrieben werden.

Digital Library Research Group

Die „Digital Library Research Group“ ist nicht wie die bisherigen Institute ein Teil der Bibliothek sondern ein Teil des „Computer Science“-Instituts. Als sol-

ches ist es mehr an den informatik- und softwaretechnologischen Aspekten von digitalen Bibliotheken interessiert. Hier werden die Grundlagen für die Erstellung von digitalen Bibliotheken geschaffen, aber bei der Entwicklung der Systeme hat dieses Institut nur beratende Funktion.

Einige Projekte der Abteilung sind:

- Dienst: Architektur und Protokoll für die Implementierung einer verteilten digitalen Bibliothek
- Fedora: „Flexible and Extensible Digital Object and Repository Architecture“. Definiert Mechanismen für die Speicherung, Manipulation, Zugangsverwaltung und Darstellung der Inhalte digitaler Bibliotheken
- D-Lib Magazine: eine Zeitschrift für Innovation und Forschung auf dem Gebiet der digitalen Bibliotheken
- Projekt Prism: ein Projekt über die Integrität von Informationen in einer verteilten Bibliothek

Cornell Institute for Digital Collections

Das Cornell Institute for Digital Collections erzeugt virtuelle Sammlungen von seltenen Materialien der Bibliotheken und Museen von Cornell.

Beispiele für Projekte des Instituts sind:

- Death of the Father: eine digitale Sammlung über das Ende von Mussolini, Hitler, Hirohito, Ceausescu, Stalin und Tito. Die Sammlung enthält Karten, Bilder, Filme und Tondokumente.
- Birds: die digitalisierte Form einer Bildersammlung von Vogeldarstellungen mit 2500 Objekten.
- Utopia: ein digitales Museum von Bildern europäischer Renaissancekunst mit 4520 Objekten.

Harvard University

Harvard wurde 1638, 16 Jahre nach Ankunft der Pilgerväter, gegründet. Harvard ist damit die älteste höhere Bildungseinrichtung in den USA. 1638 begann man mit 9 Studenten und einem einzigen Professor. Heute studieren in Harvard 18.000 Studenten in 10 Schulen. Weitere 13.000 nehmen an externen Programmen teil.

Harvards Universitätsbibliothek ist ein System von 90 Einheiten. Insgesamt verwaltet die Bibliothek 6 Millionen Bücher, 8 Millionen Mikrofilme, 500.000 Karten und 51.000 Tondokumente sowie eine große Anzahl von Objekten in diversen anderen Formaten.

In Harvard beschäftigen sich zwei Initiativen mit digitalen Bibliotheken und der Digitalisierung von Bibliotheksbeständen: die „Library Digital Initiative“ und die „Digital Imaging Group“.

Library Digital Initiative

Die Universität Harvard startete die „Library Digital Initiative“ im Juli 1998 als 5-Jahres-Programm. Ziel ist es, Möglichkeiten für die Verwaltung von digitalen Informationen zu entwickeln.

Das soll erreicht werden durch:

- den Aufbau einer technischen Infrastruktur für den Erwerb, die Verwaltung, die Auslieferung und die Archivierung von digitalem Material
- ein Team von Spezialisten zur Beratung der Bibliothekare und anderer Universitätsangehöriger
- Aufbau von Know-how
- Aufbau einer signifikanten Sammlung von digitalen Objekten

Die „Library Digital Initiative“ betreibt einen zentralen Server, in dem alle digitalen Sammlungen der Universität verwaltet werden. Die einzelnen Bibliotheken müssen dort ihre digitalen Objekte in einem von der LDI entwickelten XML-Datenformat ablegen.

Die digitalen Objekte in diesem Format werden unter anderem von der „Digital Imaging Group“ erstellt.

Digital Imaging Group

Die „Digital Imaging Group“ stellt ein Service zur Erstellung von Scans und digitalen Photographien in höchster Qualität zu Verfügung. Die erstellten Objekte sollen von permanentem Wert für Forschung und Lehre sein. Das Service steht allen Bibliotheken und Archiven von Harvard zur Verfügung.

Für die Erstellung der digitalen Objekte wurde ein digitales Photolabor nach neuesten Erkenntnissen eingerichtet. Es erfüllt dabei folgende Kriterien:

- Scannen, digitale Photographie und Image Processing-Qualitätskontrolle in einem Studio
- optimale Balance zwischen Handling und Produktion
- richtiger Umgang mit Farbe
- Konsistenz der Prozesse und Produkte

Technisch erfüllt das Labor den ISO Standard 3664 für die Darstellung von Bildern auf Monitoren. Dabei werden Umgebungslicht, Monitor, Scanner, Drucker, Kamera und sogar Farbe der Wände, Decken und Böden aufeinander abgestimmt. Vor allem die Wahl des Umgebungslichts war für das Projekt entscheidend, um die Farbe auf dem Monitor mit der Farbe auf der Vorlage vergleichen zu können.

Einen kompletten Bericht über die Ausstattung des Labors findet man im *DigiNews* Magazin vom Dezember 2000.

Trotz der hervorragenden Ausstattung des Labors erwähnen die Verantwortlichen aber immer wieder, dass die Technik allein nicht ausreicht, um gute Digitalisierungsergebnisse zu erreichen. Eine gute Organisation und qualifizierte Mitarbeiter, in diesem Fall ehemalige Profifotographen und Graphiker, sind mindestens ebenso wichtig.

METAe Projekt

- Homepage: <http://meta-e.uibk.ac.at/>

New York Public Library

- Homepage: <http://www.nypl.org/>
- Digital Library Collection: <http://digital.nypl.org/>

University of Michigan

- Homepage der Universität: <http://www.umich.edu/>
- Homepage der Bibliothek: <http://www.lib.umich.edu/>
- Digital Library Production Service: <http://www.umdl.umich.edu/>
- Digital Library Project: <http://www.si.umich.edu/UMDL/>

Cornell University

- Homepage der Universität: <http://www.cornell.edu/>
- Digitalisierungsinitiativen:
<http://www.cs.cornell.edu/prism/PrismWeb/DLCornell.htm>
- Cornell Digital Library Research Group: <http://www.cs.cornell.edu/cdlrg/>
- Cornell Institute for Digital Collections: <http://cidc.library.cornell.edu/>

Harvard

- Homepage der Universität: <http://www.harvard.edu>
- Library Digital Initiative: <http://hul.harvard.edu/ldi/index.html>
- Digital Imaging Group: <http://preserve.harvard.edu/dig/>

- DigiNews Artikel über das Harvard Fotolabor:
<http://www.rlg.org/preserv/diginews/diginews4-6.html>