

HANS PETSCHAR

## Digitale Archive und Bibliotheken

### Vorab eine Anmerkung

Bibliotheken und insbesondere Nationalbibliotheken rücken mehr und mehr ins Zentrum internationaler Digitalisierungsprojekte. Im folgenden werde ich kurz die Gründe hierfür skizzieren und die sozialen und kulturellen Konsequenzen zur Diskussion stellen. Von den Anfängen der Multimediatechnologie blieben Archive und Bibliotheken – im Unterschied zu den Museen – weitgehend verschont. Zum einen liegt dies darin begründet, daß das verstaubte Image dieser Institutionen offenbar nicht dazu angetan war, die Vertreter der Industrie dahin zu bewegen, in diese Bereiche zu investieren und prestigeträchtige, multimediale, interaktive Bibliotheks- oder Archiv-CD-ROMs oder CD-Is zu produzieren. Zum anderen waren und sind diese Institutionen hauptsächlich damit beschäftigt, ihre Grunddaten, also die Bibliothekskataloge auf eine elektronische Basis zu stellen. Die erste CD-ROM, die von der Österreichischen Nationalbibliothek Ende der 80er Jahre produziert wurde, war demzufolge eine Sammlung bibliographischer Daten. Mittlerweile hat sich die Situation verändert. Wer heutzutage die Website der Österreichischen Nationalbibliothek<sup>1</sup> besucht (Abb. 1), findet im Menüpunkt Online-Services einen Zugang zu allen Druckschriftenkatalogen der Bibliothek. Möglich wurde dies durch den 1989 gegründeten elektronischen Bibliotheksverbund, der in einem Datenbankmanagementsystem die seit 1988 erschie-

Abbildung 1: Website der Österreichischen Nationalbibliothek



nenen Hochschulschriften und die seit 1991 erschienenen Publikationen enthält. Den Inhalten dieser Datenbank steht die Fülle von etwa 6 Mio. Katalogkärtchen der älteren Druckschriftenbestände gegenüber, die in den Schlagwort- und Nominal- bzw. den Autorenkatalogen verwaltet werden. Aufgrund einer Unterstützung in Höhe von etwa 5 Mio. ATS des damaligen Bundesministeriums für Unterricht und kulturelle Angelegenheiten konnten auch diese Bestände elektronisch zugänglich gemacht werden. Im Rahmen des Projektes wurden alle Karteikarten eingescannt und eine Software entwickelt, mit der man via Internet in den Zettelkatalogen blättern kann. Blättern deshalb, weil diese elektronischen Kataloge nichts anderes sind als elektronische Abbilder der maschinengeschriebenen Karteikarten. Es gibt keinen OCR-Text der Inhalte und leider auch keine Metadaten, deren strukturierte Felder Suchabfragen nach Autor und Titel ermöglicht hätten. Der Grund hierfür liegt in den hohen Kosten. Die Österreichische Nationalbibliothek konnte keine 180 Mio. ATS bereit stellen, die die Bearbeitung und Neukatalogisierung der Altbestände, d.h. die Eingabe der Daten in Datenbankfelder nach Standardformaten, gekostet hätte. Mittlerweile sind die OCR-Softwareprodukte zwar so gut, daß relativ homogene

Printvorlagen in hoher Qualität eingescannt werden können, doch bei Katalogen, die über Jahrzehnte hinweg entstanden sind und Karteikarten, die über unterschiedliche Erscheinungsbilder verfügen, ist diese Voraussetzung zumeist gar nicht gegeben. Der Katalog 1501–1929 bildete einen Glücksfall, da dieser in den 60er Jahren überarbeitet und abgeschrieben worden ist und somit ein relativ einheitliches Erscheinungsbild aufweist. Dieser Katalog wurde daher im Rahmen unseres Projektes OCR-gelesen und in das neue Bibliotheksdatenbanksystem *Aleph 500*, das seit Anfang 1999 in Betrieb ist, integriert. Dennoch mußten noch ein paar tausend Korrekturen durchgeführt werden, die durch das OCR-Lesen entstanden sind. Der Versuch der Digitalisierung unserer Katalogbestände zeigt recht gut die Kehrseite der digitalen Medaille, denn so sieht in Wahrheit die Realität für alle großen Archive und Bibliotheken aus. Von digitaler Euphorie kann demnach keine Rede sein. Die Mehrzahl der Informationen, Quellen und vor allem der Organisationsstrukturen in Europa sind keineswegs elektronisch. Und die Kosten, die nur in der Bereitstellung der digitalen Grunddaten liegen, sind exorbitant hoch. Dennoch zeigt ein internationaler Vergleich, daß vor allem die großen Nationalbibliotheken seit Jahrzehnten sehr viel in die Schaffung von digitalen Grundkatalogisaten investiert haben.

## **Internationale Erfahrungen**

---

Wer in der Library of Congress oder seit neuestem in der Bibliothèque nationale de France recherchieren will, kann dies nach Belieben, weil die Druckschriftenbestände sämtlich in standardisierten Katalogisaten elektronisch erschlossen sind. Millionen an Francs und Dollars wurden bereits vor Jahren in die Schaffung dieser Basisdaten investiert. Heute geht es in diesen Ländern bereits um die Katalogisierung von Objekten aus den Spezialsammlungen und um den Aufbau von multimedialen Datenbanken und deren Vernetzung. Im folgenden seien aus der Fülle an Bibliotheksprojekten einige hervorgehoben.<sup>2</sup>

### **Library of Congress**

Die Schaffung einer breiten Grundlage von digitalen Objekten wird als nationale Aufgabe angesehen. Die Library of Congress<sup>3</sup> plant

die Digitalisierung von 5 Mio. Americana bis zum Jahr 2000 innerhalb des nationalen Digitalisierungsprogrammes *American Memory*. Dieses Prestigeprojekt wird vom Kongreß mit 15 Mio. USD und durch private Sponsoren mit 45 Mio. USD gefördert.

### **Bibliothèque Nationale de France**

Die Bibliothèque Nationale de France<sup>4</sup> verfolgt mit dem Projekt *Gallica* ein ähnlich ambitioniertes Ziel, welches die Digitalisierung von 300.000 Bildern und 50.000 Büchern bis Ende 2000 vorsieht. Inhaltlich will man sich auf Bestände aus den Sammlungen der Bibliothèque Nationale und auf unikale Objekte konzentrieren. Da insbesondere beim Aufbau von großen digitalen Bildsammlungen die Katalogisierung der Objekte der teuerste und zeitaufwendigste Aspekt ist, sollen relativ geschlossene Sammlungen, die bereits erschlossen sind, bevorzugt werden. Besondere Bedeutung wird an der Bibliothèque Nationale de France auch dem editorischen Aspekt beim Aufbau von digitalen Kollektionen beigemessen, der sich durch den Auswahlcharakter dieser Sammlungen und deren Aufbereitung für die Netzwerkkommunikation zwangsläufig ergibt.

### **British Library**

Die British Library<sup>5</sup> startete mit ihrem *Initiatives for Access*-Programm eine Reihe von Projekten, die unterschiedliche Aspekte von digitalen Bibliotheken zum Gegenstand haben. Die British Library setzt auch in diesem Bereich auf ihre Stärken wie Patente, Document Delivery und Kooperation mit Partnern aus der Wirtschaft. Im Pilotprojekt *PIX* wurde die digitale Bildarchivierung mit einem Sample von 13.000 Bildern erprobt. Die Digitalisierung erfolgte auf Photo-CD-Basis; für das Retrieval wurde ein Bildsuchprogramm entwickelt. Das Projekt wurde erfolgreich abgeschlossen, wird aber vorerst nicht fortgesetzt. Insbesondere die Katalogisierung der Objekte stellte eine entscheidende zeitliche Hürde dar und ist im großen Ausmaß nicht zu bewerkstelligen. Zudem ist die Software inzwischen veraltet und stellt eine Einzellösung dar, die nicht in das Bibliothekssystem integrierbar ist. Die British Library nimmt gegenüber Großprojekten mittlerweile eine vorsichtige Haltung ein. Bevorzugt werden Einzelprojekte bzw. Lösungen konkreter Probleme wie die Entwicklung einer Hypermedia-Software zur Darstellung und Analyse des Beowulf-Manuskriptes oder die Er-

probung hochauflösender Digitalisierungstechniken für die *Treasures* der Bibliothek.

## Dänische Nationalbibliothek

Die Dänische Nationalbibliothek<sup>6</sup> ist mit einem Budget von 212 Mio. DK federführend beteiligt am Aufbau eines digital library Projektes. Hierbei werden neben Sondersammlungen des Bildarchivs auch rund 100.000 dänische Portraits digitalisiert. Die Katalogisierung erfolgt in einer Groberfassung. Durch eine grundlegende Strategiestudie vorbereitet, wird das dänische Digitalisierungsprojekt von drei Ministerien unterstützt. Zielsetzung ist es, ein vergleichbar kleines Land mit einer historischen Nationalbibliothek an großen internationalen Digitalisierungsprogrammen zu beteiligen.<sup>7</sup>

Zusammenfassend ergibt sich folgendes Bild: Bis zum Jahr 2000 werden weltweit von Nationalbibliotheken Projekte zum Aufbau nationaler digitaler Sammlungen durchgeführt worden sein. Der Aufbau einer »kritischen Masse« digitaler Objekte wird als strategisch wichtige Aufgabe für die Zukunft der Bibliotheken angesehen. Die digitalen Sammlungen stellen eine Verbesserung des Zugangs zu den Beständen dar und ergänzen die Serviceleistungen der Bibliotheken. Bei einer genaueren Betrachtung lassen sich aber doch nach den Erfahrungen aus den ersten zehn Jahren mit großen Digitalisierungsprogrammen entscheidende Kursänderungen und neue strategische Entwicklungslinien für den Aufbau digitaler Bibliotheken feststellen. Die entscheidenden Punkte sollen an dieser Stelle kurz skizziert werden. Zu Beginn der 90er Jahre wurde unter dem Stichwort der *digitalen Bibliothek* bzw. des *digitalen Archivs* in erster Linie eine möglichst großflächige Digitalisierung von Objekten verstanden. Als Beispiel sei die *Biblioteca electronica del Archivo General de Indias in Sevilla* genannt. Zwischen 1985 und 1992 wurde eines der ambitioniertesten Digitalisierungsprojekte, die es bislang gab, durchgeführt. Aus einem Gesamtbestand von 90 Mio. Seiten, denen 45 Mio. Dokumente zugrunde lagen, sollten 10 Prozent digitalisiert und zugänglich gemacht werden. Das Projekt beinhaltete den Aufbau einer Bilddatenbank, den Aufbau einer bibliographischen Datenbank und die Einrichtung eines Benutzungs-systems. Das Sevillaprojekt, 1992 im Rahmen der Weltausstellung groß präsentiert, wurde vielfach als modellhaftes Beispiel für die Zusammenarbeit von Technikern und Archivaren genannt und er-

hielt hierfür 1994 den Erasmuspreis. Seither ist es aber relativ still geworden um dieses erste große Digitalisierungsprojekt. Eine Benutzung der Datenbank ist nur vor Ort möglich. Eine Kommunikation über Internet, heutzutage Kennzeichen aller großen Projekte ist, wurde offenbar nicht in die damalige Planung miteinbezogen. Zudem haben sich bis heute die Speichertechnologien, vor allem im Bereich der Massenspeicher und der Langzeitarchivierung, rapide weiterentwickelt. Was den alten – und zehn Jahre ist bereits uralt – digitalen Sammlungen geblieben ist, sind eine Unmenge von Daten. Und die schmerzliche Erfahrung, daß die Kosten sowohl für eine Migration, einen Datentransfer als auch die Kosten für die Aufrechterhaltung und Wartung des Systems gewaltig unterschätzt wurden. Inwieweit einige dieser Faktoren auch auf das Archivo General de Indias in Sevilla zutreffen, wird sich wohl nur durch einen realen Besuch und ein ehrliches Gespräch mit den verantwortlichen Archivaren klären lassen. In den allgemeinen Projektbeschreibungen finden sich derlei Dinge naturgemäß nicht. Eine ausführliche Darstellung der Strategie, der Ergebnisse des Projektes und der weiteren Ausbaupläne liefert jedoch eine jüngst vom Council on Library and Information Resources publizierte Studie von Pedro Gonzáles (Gonzáles 1998). Generell läßt sich feststellen, daß vor allem in den ersten Projekten die digitale Bibliothek rein als technisches Problem gesehen wurde und die strukturellen, organisatorischen und sozialen Konsequenzen der Digitalisierung zu wenig Beachtung fanden. So vollzog sich der Aufbau der digitalen Kollektionen weitgehend unabhängig von den vorhandenen Organisationsstrukturen.

## Die Digitale Bibliothek

---

Was ist eigentlich eine *Digitale Bibliothek* und wie definiert man diese? Eine mögliche Definition liefert die 1995 in den USA gegründete Digital Library Federation:

*Digital libraries are organizations that provide the resources, including the specialized staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily and economically available for use by a defined community or set of communities (Water 1998).<sup>8</sup>*

Diese wunderbare Definition hat nur einen kleinen Schönheitsfehler: sie ist im Präsens formuliert und spricht doch von der (vielleicht noch fernen) Zukunft. Denn würde man alle Punkte, die hier formuliert sind, ernst nehmen und als definitorische Voraussetzung einer Digitalen Bibliothek ansehen, so müßte man feststellen, daß es auf der ganzen Welt noch keine gibt. Meiner Ansicht nach gibt es gegenwärtig nur Experimente und Programmatiken. Dies festzustellen und zu erkennen ist für diejenigen, die nicht mit der medialen Verbreitung, sondern mit der Umsetzung und der Planung betraut sind, nichts Negatives. Im Gegenteil, es ist eher der erste Schritt zu einer möglichen Realisierung. Daher ist die programmatische Definition einer Digitalen Bibliothek, wie sie von der Digital Library Federation gegeben wurde, unbedingt als Aufforderung zu verstehen. Als eine Aufforderung und möglicherweise auch Warnung, in der Planung für die Digitalisierung alle Voraussetzungen und auch die Konsequenzen für den Aufbau, die Archivierung und die Benützung digitaler Daten zu berücksichtigen. Länder wie Österreich, aber auch Deutschland, die mit ihren nationalen Digitalisierungsprogrammen noch am Anfang stehen, haben den Vorteil, daß sie hier auf konkrete Erfahrungen und Empfehlungen anderer Bibliotheken aufbauen können. So muß der Entwurf von strategischen Entwicklungsplänen unbedingt als notwendige Voraussetzung für den Erfolg nationaler Programme verstanden werden. Ohne eine nationale Strategie zum Gebrauch der Informations- und Kommunikationstechnologie auf dem Sektor des kulturellen Erbes treten mannigfaltige Probleme auf, wie Seamus Ross und Maria Economou in einer britischen Studie darlegten:

*It is unlikely that there will be national or international comparability between digital resources. Priorities for creation of digital materials will not be coherent or complementary between heritage institutions. Substantial sums will continue to be spent each year developing and curating digital resources without a clear understanding whether this is being done to best advantage or practise, and finally, it will be impossible for funding agencies to prioritise the relative merits of different proposals coming forward for support (Ross/Economou 1998).*

## Resümee

---

Seit den 90er Jahren hat die digitale Kommunikationstechnologie endgültig Eingang in den Alltag von Bibliotheken und Archiven gefunden. Die Digitalisierung von Objekten erlaubt es, die Möglichkeiten der digitalen Netzwerkkommunikation auszuschöpfen und einen verbesserten Zugang zu den Objekten zu schaffen. Vervielfältigung ohne Zerstörung (oder Beeinträchtigung) der Objekte, schnelle Datentransferierung über Netzwerke und die Möglichkeit zur Integration in andere Dokumente sind vorteilhafte Aspekte der Digitalisierung, die außerdem Bibliotheken auch neue Serviceleistungen und Dienstleistungsangebote ermöglichen. Demgegenüber steht die Tatsache, daß Institutionen, die jahrhundertlang die Garanten für Stabilität und Kontinuität waren, plötzlich mit ihren digitalen Objekten und deren Kommunikation, d.h. also mit ihren Quellen und mit ihrer ganzen Organisationsstruktur, auf Gedeih und Verderb der Kommunikationstechnologie mit ihren rasanten Veränderungen ausgeliefert scheinen. Durch die Digitalisierung ihrer Objekte werden Bibliotheken und Archive zu *Publishern* und *Content Providern*. Dies bedeutet einen verstärkten Aufbau von neuen Organisationsstrukturen innerhalb der Institutionen selbst und eine Neuverteilung der Rollen auf dem ganzen Publikationsmarkt. Der Argwohn, mit dem professionelle Informationsanbieter zur Zeit auf die Informationsproduzenten aus dem öffentlich-rechtlichen Bereich schauen, ist verständlich. Weniger verständlich hingegen ist, daß von den Institutionen viel zu wenig in die andere, die marktwirtschaftliche Richtung geschaut wird.

Eine sinnvolle Digitalisierung von Objekten bedeutet in erster Linie die Hinterfragung und die Neudefinition von Selektionskriterien als auch die Berücksichtigung von hierarchisch gewachsenen Organisationsstrukturen, rechtlichen Aspekten und ökonomischen Faktoren. Die Kosten sind nach wie vor hoch, denn sie betreffen neben der reinen Digitalisierung der Objekte auch die Kosten für die Strukturierung von Informationen, die Schaffung von Zugangsmöglichkeiten, ebenso wie Betriebskosten und die Kosten für den Datentransfer. All dies sind entscheidende Faktoren für den Erfolg eines Projektes, und sie müssen unbedingt in der Planung berücksichtigt werden. Nicht zuletzt aus diesen Gründen ist der internationale Informationsaustausch bei der Planung und Realisierung nationaler Prestigevorhaben unumgänglich. Die digitale Euphorie ist keine *business strategy*. Nur eine vorausschauende Pla-



nung trägt in diesen »euphorischen Zeiten« dazu bei, das kulturelle Erbe auch nachfolgenden Generationen zugänglich zu machen.

---

### Anmerkungen

- 1 Österreichische Nationalbibliothek: [www.onb.ac.at](http://www.onb.ac.at).
- 2 Unter <http://mpiz-koeln.mpg.de> der Max-Planck-Gesellschaft findet sich eine Link-Liste mit weiterführenden URLs von Bibliotheksdatenbanken.
- 3 Library of Congress: <http://lcweb.loc.gov>.
- 4 Bibliothèque Nationale de France: <http://www.bnf.fr>.
- 5 British Library: <http://portico.bl.uk>.
- 6 Het kongelige Bibliotek: <http://kb.bib.dk>.
- 7 Interessant in diesem Zusammenhang auch Danmarks Elektroniske Forskningsbibliotek: <http://www.detlink.dk>.
- 8 <http://www.clir.org/pubs/issues/issueso4.htm>.

### Literatur

- Gonzáles, Pedro (1998): »Computerization of the Archivo General de Indias: Strategies and Results«. In: Council on Library and Information Resources, Washington D.C.
- Ross, Seamus/Economou, Maria (1998): »Information and Communications Technology in the Cultural Sector. The Need for National Strategies«. D-Lib Magazine, June.
- Water, Donald J. (1998): »What are Digital Libraries«. CLIR issues, 4.

