



IFLA
2005
OSLO

World Library and Information Congress: 71th IFLA General Conference and Council

"Libraries - A voyage of discovery"

August 14th - 18th 2005, Oslo, Norway

Conference Programme:

<http://www.ifla.org/IV/ifla71/Programme.htm>

2nd Version: Juni 21, 2005

Code Number:

034-G

Meeting:

150 SI - ICABS (IFLA/CDNL Alliance for Bibliographic Standards)

Nestor and kopal - zwei nationale Initiativen zur Sicherstellung von Langzeitverfügbarkeit digitaler Dokumente in Deutschland

Reinhard Altenhöner

Leiter der Abteilung IT
Die Deutsche Bibliothek

Abstract

Der Umbruch des Publikationsmarktes hin zur elektronischen Publikation in Netzen verändert das Aufgabenprofil der Bibliotheken – und insbesondere der Bibliotheken, die für die Überlieferung des kulturellen Erbes einer Nation verantwortlich sind – erheblich. Diese Ausgangssituation stellt sich in Deutschland insofern in spezifischer Weise, als hier eine große Zahl unterschiedlicher Bibliotheken mit der Aufgabe der Sammlung, Erschließung und Archivierung von Publikationen betraut sind. Dies ist eine Aufgabe, die es nun für eine veränderte Publikationswelt unter organisatorischen und technischen Gesichtspunkten neu zu formulieren gilt. Dabei besteht über alle Strukturen hinweg Klarheit darüber, dass diese Aktivitäten in kooperativer Form erfolgen sollen.

In diesem Beitrag werden daher aktuelle Ansätze in Deutschland vorgestellt. Während einerseits auf eine Klärung der rechtlichen Rahmenbedingungen für das neue Aufgabenspektrum hin gearbeitet wird, wurden auf der anderen Seite zwei Initiativen ins Leben gerufen: Einmal der Aufbau eines institutionenübergreifenden Kompetenznetzwerks für Fragen der Langzeitarchivierung (nestor), auf der anderen Seite die kooperative Entwicklung eines Depotsystems für digitale Ressourcen (kopal). Beide Aktionslinien, an denen Die Deutsche Bibliothek federführend beteiligt ist, werden präsentiert.

1. Allgemeine Sicht

Das Publizieren erfolgt heute zunehmend elektronisch. Allerdings wissen wir noch immer nicht, wie wir elektronische Veröffentlichungen, Kunstwerke, Bild- und Klangdokumente, Akten, Primärdaten und Musik so archivieren können, dass sie dauerhaft lesbar und somit auch in Zukunft für alle zugänglich und benutzbar sind. Was wir sagen können ist: Welche Strategie auch immer in Zukunft verfolgt wird, um digitale Inhalte zur Verfügung zu stellen, die Benutzbarkeit wird von dem Vorhandensein eines Bit-Streams (Datenstroms) abhängen, dessen Datenintegrität und Authentizität über die Jahre erhalten wurde. Nur auf der Grundlage dieses Bit-Streams wird maschinell die Dokumentenwiedergabe beim zukünftigen Zugriff auf digitale Objekte möglich. Verschiedene Strategien stehen zur Diskussion, die in zwei Wirkungslinien zusammengefasst werden können:

- die elektronischen Objekte in einer kontrollierten Umgebung zu migrieren

oder

- die historische Systemumgebung der Ursprungsform des Objekts zu emulieren – einschließlich der Emulation von Hardware und Systemsoftware.

Neben der Verfügbarkeit des Bit-Streams selbst benötigen wir zusätzlich Metadateninformationen, die speziell die technische Information zu den gespeicherten Objekten betreffen, ihre ursprüngliche technische Umgebung und ihre Speichergeschichte. Die Aufgabe, ein System und eine Infrastruktur für eine Langzeitarchivierung aufzubauen, ist sehr komplex und erfordert viele Geldmittel und personelle Ressourcen; ferner benötigen wir ein Wissensnetzwerk und den Austausch von Fachwissen – sowohl auf nationaler wie internationaler Ebene.

2. Sicht Der Deutschen Bibliothek

Die Deutsche Bibliothek¹ ist die Nationalbibliothek und das nationalbibliografische Informationszentrum der Bundesrepublik Deutschland. Die Bibliothek ist zuständig für die Sammlung, Bearbeitung und die bibliografische Verzeichnung aller deutschen und deutschsprachigen Publikationen, die seit 1913 veröffentlicht wurden. Diese Aufgabe gründet auf dem gesetzlichen Sammelauftrag, der die bibliografische Verzeichnung und Langzeiterhaltung aller Publikationen, die in Deutschland herausgegeben werden oder im Ausland in deutscher Sprache erscheinen, beschreibt. Das Gesetz umfasst auch digitale Publikationen, die auf physikalischen Trägern verbreitet werden, es bezieht sich aber bis heute nicht auf Online-Publikationen. Zurzeit warten wir auf eine neue Gesetzgebung, die unseren Verantwortungsbereich auf alle Arten von Netz- oder elektronischen Publikationen ausdehnen wird. Wir erwarten, dass als Folge des neuen Gesetzes viele regionale Pflichtexemplarbibliotheken für ihre Sammlungen digitaler Publikationen eine neue Rechtsgrundlage erhalten.

Auf diese Situation müssen wir vorbereitet sein. Daher hat Die Deutsche Bibliothek im Laufe der vergangenen Jahre einige Initiativen gestartet, die die Langzeitarchivierung digitaler Publikationen in Deutschland fördern. In Vorgesprächen mit Verlegern, Bibliotheksexperten, IT-Spezialisten und Regierungsbeamten wurden Grundlagen für die Sammlung von Online-Publikationen definiert und in einem Dokument, das durch das Verlegerkomitee des Börsenvereins des Deutschen Buchhandels im Juni 1997 verabschiedet wurde, zusammengefasst:

- Alle Online-Publikationen müssen auf Anfrage über Netzwerke oder auf physischen Datenträgern abgeliefert werden.
- Online-Publikationen, die in verschiedenen Formen zugänglich sind, müssen in dem von der Bibliothek geforderten Format abgeliefert werden.

¹ Die Deutsche Bibliothek. <http://www.ddb.de/>

- Publikationen identischen Inhalts, die sowohl auf physischen Datenträgern wie auch als Online-Publikation vertrieben werden, müssen in beiden Publikationsformen abgeliefert werden.
- Online-Publikationen identischen Inhalts, die gleichzeitig von mehreren Anbietern vertrieben werden, müssen nur einmal abgeliefert werden.

Gemäß dieser Richtlinie hat Die Deutsche Bibliothek zusammen mit Verlegern und Herstellern in einer mehrjährigen Testphase Abgabeverfahren sowie Sammelverfahren und Verfahren zur Langzeitarchivierung von Online-Publikationen getestet. Im Verlauf dieser Testphase hat die Arbeitsgruppe 'Elektronische Depot-Bibliothek', die für Die Deutsche Bibliothek erforderlichen Voraussetzungen, eine Depot-Bibliothek für Online-Publikationen zu werden, untersucht und geschaffen.

Seit 1998 werden Online-Dissertationen und Hochschulschriften (36.000 bis heute) gesammelt, archiviert und auf einem Dokumentenserver zur Verfügung gestellt. Seit 2000 werden elektronische Zeitschriften gesammelt, und seit 2001 betreibt Die Deutsche Bibliothek eine Abgabeschnittstelle für Online-Publikationen. Während des Abgabeverfahrens hat Die Deutsche Bibliothek auch technische Metadaten (an)gefordert, die für Aufbewahrungszwecke relevant sind. Zwischen der Arbeitslast, die die Verleger unter der Bedingung freiwilliger Abgabe bereit sind, auf sich zu nehmen, und den aufwändigen Anforderungen zukünftiger Archivierungsprozesse im Depotsystem muss ein Kompromiss geschlossen werden. Ferner war Die Deutsche Bibliothek Mitglied des Europäischen Nedlib-Projektes, das das OAIS-Modell übernommen und Vorschläge zum Arbeitsablauf für die Integration von Bibliotheksabläufen entwickelt hat.

Andere wichtige Erfahrungen:

- Die Deutsche Bibliothek hat mit ihrem Multimedia-Bereitstellungssystem (MMB) Erfahrungen gesammelt. MMB ermöglicht die Speicherung und den Zugriff auf digitale Objekte auf physischen Datenträgern. Es wurde für verschiedene Objekttypen implementiert (workstation image, application installation kit, file collection, presentation object), um die Wiedergabe komplexer digitaler Objekte zu ermöglichen.
- Eine andere Aktivität beinhaltet die Entwicklung einer Persistent-Identifizier-Infrastruktur für Deutschland. Die Anwendung von Persistent Identifiern garantiert die Stabilität und eine eindeutige Möglichkeit, ein digitales Objekt im Netz² zu adressieren.

3. Deutschlands föderale Struktur

Deutschland besitzt eine föderale Struktur mit wichtigen Elementen der Selbstverwaltung in den einzelnen Ländern, die besonders das Bildungssystem sowie Wissenschaft und Forschung betreffen. Die Existenz vieler Regionalbibliotheken mit Pflichtabgaberecht für ihre jeweilige Region ist ein weiterer Bestandteil der föderalen Struktur. In Anbetracht der Bedeutung der Langzeitarchivierung innerhalb der föderalen Struktur Deutschlands ist es offensichtlich, dass eine erfolgreiche Lösung nur durch einen kooperativen Ansatz erreicht werden kann. Dies betrifft in erster Linie den organisatorischen Aspekt, doch gibt es viele praktische und technische Gründe, warum versucht wird, die Verantwortung, elektronische Objekte im Netz zu sammeln, zu verteilen.

Deshalb muss die Auswahl der Partner, die wir zur Implementierung einer Langzeitarchivierungsstrategie und -infrastruktur in Deutschland brauchen, sowohl die organisatorische wie auch die betriebliche Ebene abdecken.

² So im Projekt "EPICUR - Enhancement of Persistent Identifier Services - Comprehensive Method for unequivocal Resource Identification" entwickelt, bekommen die persistent identifier eine konkrete Aufgabe im Kontext der Metadatenentwicklung für elektronische Publikationen.

4. Nationale Initiativen in Deutschland

Es gibt in Deutschland zwei Initiativen, die sich mit dem Problem der Langzeitarchivierung befassen:

- Auf der organisatorischen Seite gibt es **nestor** (Network of Expertise in Long-Term Storage of Digital Resources), das eine Kompetenzplattform für den Austausch von Kenntnissen und Erfahrungen im Bereich der Langzeitarchivierung und den Austausch von Experten und Fachwissen zwischen unterschiedlichen Arten von Institutionen des Kulturerbes aufbauen soll.
- Auf der praktischen Seite wurde zum Aufbau eines auf OAIS basierenden Langzeitarchivierungssystems das Projekt **kopal** (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen) eingerichtet.

5. nestor

Das vom Bundesministerium für Bildung und Forschung finanzierte Projekt nestor, die „Allianz für Deutschlands digitales Gedächtnis“, läuft seit 2003 und wird 2006 beendet sein. Unter Federführung Der Deutschen Bibliothek gibt es mehrere Partner aus dem Bibliotheksbereich (Bayerische Staatsbibliothek, Niedersächsische Staats- und Universitätsbibliothek), ein Medienzentrum (Computer- und Medienservice der Humboldt-Universität zu Berlin), Partner aus Archiven (Generaldirektion der Staatlichen Archive Bayerns) und aus dem Museumsbereich (als Multiplikator das Institut für Museumskunde, Berlin). Zusätzlich sind im nestor-Beirat Verleger, Repräsentanten aus Wissenschaft und Technik, Museen, Archive, Bibliotheken und Universitäten und schließlich Mitglieder aus Kultur und Politik sowie Forschungsinstitute / Computerzentren vertreten.

Ziel von nestor ist es, vorhandenes Wissen, Menschen und Fachkenntnisse zur Langzeitarchivierung digitaler Quellen als Ausgangspunkt für ein zukünftiges Bündnis des digitalen Gedächtnisses Deutschlands zusammenzubringen.

Deshalb wollen wir im Projekt:

- Ein Netzwerk für Information und Kommunikation über aktuelle und zukünftige Langzeitarchivierungs-Aktivitäten in Deutschland schaffen
- Eine bereichsübergreifende Gemeinschaft zur Förderung und Unterstützung von Langzeitarchivierungs-Aktivitäten gründen und ein Bewusstsein dafür in der Gesellschaft wecken
- Ein Zusammenwirken gegenwärtiger Aktivitäten in Deutschland und eine Kooperation mit internationalen Partnern und Projekten bewirken
- Strategien für die Koordination von Langzeitarchivierungs-Aktivitäten in Deutschland entwickeln
- Eine dauerhafte Organisationsform vorbereiten, die nach Ende des Projektes 2006 den entstandenen Dienst eines Wissens-Netzwerks fortführt.

Die Aufgaben sind im Einzelnen:

- Informationen zum Thema sammeln und bereitstellen
- Kompetenzen bündeln, sichtbar und verfügbar machen
- Kooperationen fördern und eine gemeinsame Problemlösung unterstützen
- Materialien zu technischen, organisatorischen und rechtlichen Aspekten erarbeiten
- Modelle präsentieren, zur Diskussion stellen und die breite Nachnutzung guter Verfahren anregen
- Standards fördern und Deutschland in internationalen Standardisierungsgremien vertreten

- Sammelrichtlinien und Auswahlverfahren für die Archivierung digitaler Quellen erarbeiten
- Kriterien für vertrauenswürdige digitale Archive sowie Zertifizierungsverfahren für Archivserver entwickeln
- die Übernahme von Daueraufgaben abstimmen
- ein größeres Problembewusstsein sowohl in der Dokumentations-Fachwelt als auch in der Öffentlichkeit schaffen
- Vorbereitung einer Organisationsform, welche die Belange der Langzeitarchivierung digitaler Ressourcen dauerhaft koordiniert und vertritt.

Ein wichtiger Teil (um die Aktivitäten zu veranschaulichen) sind Arbeitsgruppen und fachliche Expertisen. Zurzeit gibt es folgende Gruppen:

- nestor Arbeitsgruppe Vertrauenswürdige Archive
- nestor Arbeitsgruppe Multimediaarchivierung
- nestor Arbeitsgruppe Archivierungsregeln und –verfahren (preservation policies), Auswahlkriterien (in Vorbereitung)

Die Expertenberichte behandeln folgende Themen:

1. Langzeitarchivierung von eJournals
2. Bedarfsanalyse der Archivierung multimedialer Inhalte
3. Entwicklung eines Beschreibungsprofils für eine nationale Langzeitarchivierungs-Strategie (Preservation Policy)
4. Digitale Langzeitarchivierung und Recht
5. Studie zum Stand vorhandener Forschungsdaten und Rohdaten aus wissenschaftlichen Tätigkeiten
6. Vergleich bestehender Archivierungssysteme
7. Digitalisierung und Erhalt von Digitalisaten in deutschen Museen

Einige dieser Berichte wurden bereits veröffentlicht und stehen im Netz zur Verfügung.

6. kopal

Ausgehend von einem eher praktischen Ansatz geht kopal das Problem der Langzeitarchivierung technisch wie organisatorisch kooperativ an: Finanziell unterstützt durch das Bundesministerium für Bildung und Wissenschaft, soll kopal eine innovative technische Lösung in Form eines nachnutzbaren Langzeitarchivs für digitale Daten entwickeln. Die Lösung basiert auf DIAS (Digital Information and Archiving System), das gemeinsam von IBM und der Nationalbibliothek der Niederlande in Den Haag entwickelt wurde.

Innerhalb des im Juli 2004 gestarteten Projekts werden digitale Materialien aller Art der Partner Die Deutsche Bibliothek und Niedersächsische Staats- und Universitätsbibliothek (SUB Göttingen) im Massenverfahren in das Langzeitarchiv eingestellt. Sie sollen von digitalen Dokumenten in Form von PDF, TIFF oder TeX bis hin zu komplexen Objekten wie digitalen Videos reichen.

Das System wird nach internationalen Standards im Bereich der Langzeitarchivierung und Metadaten im Rahmen des OAIS-Frameworks (Open Archival Information System) implementiert. Eine transparente Integration in vorhandene Bibliotheks- und Informationssysteme ist dabei gewährleistet und ein wesentliches Ziel des Projekts.

Um eine offene, ausbaufähige und anpassungsfähige Lösung zu erhalten, haben wir Partner aus ganz unterschiedlichen Bereichen in das Projekt integriert. Ausgehend von den

unterschiedlichen Absichten der Partner, wurde entschieden, den Aspekt des Betriebs von dem der Entwicklung zu trennen. Zum Beispiel möchte unser Partner für den Betrieb des Systems (GWGD) Erfahrungen mit einem gut dokumentierten und skalierbaren System sammeln, da er weitere Partner gewinnen möchte, die die Vorteile des Systems für den eigenen Bedarf nutzen werden. Andererseits gibt es eine neutrale Plattform, auf der die Entwicklungspartner (Die Deutsche Bibliothek und SUB Göttingen) gemeinsame Lösungen für die Software und die Erhaltungsverfahren finden werden. Zusätzlich muss das System in der Lage sein, eine große Anzahl von Benutzern so zu bedienen, dass es die Unabhängigkeit der einzelnen Partner garantiert.

Hinsichtlich der Softwarearchitektur wurde zwischen der Kernfunktionalität des Archivsystems und den Umgebungstools, die die Homogenisierung und den Transfer digitaler Objekte in das System abwickeln, unterschieden. Die Abtrennung einer Kernfunktionalität (DIAS-Core) erfordert genau definierte Schnittstellen und eine offene Nutzbarkeit für zukünftige Partner: Die Werkzeuge zum Aufbau standardisierter Submission Information Packages (SIPs) und deren Import ins System werden von den Partnern unter einer Open-Source-Lizenz (GPL) entwickelt. Für die Entwicklung des Präsentationssystems für Objekte aus dem Archivsystem verfahren wir nach den gleichen Regeln: Wir möchten von den Einzelfalllösungen unabhängig werden und uns stattdessen auf genau definierte Schnittstellen beziehen, die für andere Partner und Systeme offen sind.

Der Entwicklungspartner für die Kernfunktionalität (DIAS Core) ist die Firma IBM Deutschland GmbH. Dies ermöglicht eine professionelle Anpassung der Softwarekomponenten und bietet langfristige Unterstützung.

kopal integriert so von Anfang an mehrere unterschiedliche Partner an verschiedenen Standorten. Das Langzeitarchiv für digitale Informationen bietet damit zukünftig auch anderen Einrichtungen die Möglichkeit, langfristig über ihre Daten zu verfügen. Somit erlaubt kopal Nutzungsmöglichkeiten über Bibliotheken hinaus auch für Wissenschaft, Wirtschaft und Verwaltung.

Wichtige ergänzende Bestandteile zum bestehenden DIAS-System sind:

- Realisierung von Überwachungs- und Monitoring-Funktionen zur Vorbereitung der Langzeiterhaltung digitaler Dokumente
- flexible Datenim- und -exportfunktionen, die auf der Grundlage des Objektbeschreibungsschemas METS (Metadata Encoding and Transmission Standard of the Library of Congress, USA) und von LMER (Langzeitarchivierungsmetadaten für elektronische Ressourcen) gründen³
- Nachnutzbarkeit durch weitere Institutionen und Unternehmen.

Aus diesen Gründen wird viel Mühe darauf verwendet, ein leistungsstarkes Datenmodell und eine leistungsfähige Architektur innerhalb des Systems zu definieren. Für den gesamten Projektverlauf sind verschiedene Informationsveranstaltungen sowie Workshops im nationalen wie internationalen Kontext geplant.

³ <http://www.loc.gov/standards/mets/> und <http://www.ddb.de/standards/pdf/lmer12.pdf>