

## **Langzeitarchivierung in Bibliotheken - Zum Stand der Entwicklung insbesondere in Deutschland**

Elmar Mittler

Die im dritten Jahrhundert vor Christus von Ptolemäus II. gegründete Bibliothek von Alexandria hat in langer Aufbauarbeit erreicht, praktisch alle relevanten Texte der Antike (einschließlich des Alten Testaments) in ihren Papyrusrollen in textlich überprüften Ausgaben auf Papyros zu speichern. Oft wird der Verlust dieser Bibliothek auf einen Brand zu Zeiten Cäsars zurückgeführt. Doch es gibt viele Gründe diese Version des Untergangs zu bezweifeln. Es war nämlich gar nicht nötig, die Bibliothek mit Gewalt zu zerstören: da Papyrus im Klima Alexandriens höchstens 100 bis 150 Jahre überdauerte, war es notwendig, die Texte kontinuierlich abzuschreiben, um ihre Bereitstellung zu sichern. Es genügte also, dass die bibliothekarische Infrastruktur nicht auf dem erforderlich hohen Niveau gehalten werden konnte, um die Bibliothek schleichend zu verlieren.

In einer ähnlichen Situation befinden wir uns im Zeitalter digitaler Information. Elektronische Medien sind flüchtig. Dabei beschränkt sich die langfristige Archivierung digitaler Ressourcen nicht auf das regelmäßige Erstellen von Backups der Daten. Die Sicherung des „Datenstroms“ der digitalen Ressourcen reicht für Sicherstellung der langfristigen Benutzbarkeit noch nicht aus, denn zusätzlich sind die für die Nutzung erforderlichen Hardware- und Softwareumgebungen beständigem Wandel unterworfen.

Die Digitale Langzeitarchivierung erweist sich damit als Herausforderung bisher nicht gekannter Qualität für Bibliotheken und alle anderen „heritage institutions“.

Doch das Problem ist in Wirklichkeit noch viel größer: es sind auch die Quantitäten an Dokumenten, die weit über das bekannte Maß hinausgehen; außerdem bestehen sie aus sehr unterschiedlichen Einheiten. Elektronische Publikationen (z.B. elektronische Zeitschriften, e-Books) können offline oder online sein und als Mischformen auftreten. Während analoge Dokumente immer statisch sind, ist es leicht, elektronische Materialien in immer wieder aktualisierten Formen als dynamische Dokumente zu erstellen. Während man bei gedruckten Materialien sich auf relativ klare Regeln des Publizierens verlassen kann, gibt es im elektronischen Bereich eine fließende Grenze zwischen Kommunikation und Publikation. Die Zahl der Personen und Einrichtungen, die elektronisch publizieren (können), ist fast ins Unübersehbare gestiegen, weil im Prinzip jeder im Internet veröffentlichen kann. Die Verbreitung findet auf kaum noch kontrollierbaren unterschiedlichen Kanälen statt. Zu den gedruckten Texten, die bestenfalls durch Bilder und Tabellen ergänzt wurden, treten multimediale Publikationen, Audio- und Videomaterialien oder Kombinationen von allen Möglichkeiten in virtuellen Lehr- und Lernobjekten. Zu den bisher in getrennten analogen Medien verbreiteten Materialien treten auch neue Inhalte in das Blickfeld. Wissenschaftliche Primärdaten von Messungen oder statistische Daten größeren Umfangs sollen zusätzlich zu den Texten bereitgestellt und dauerhaft zugänglich gehalten werden.

Schnelle und kostengünstige Verbreitbarkeit und Flexibilität des Einsatzes digitaler Publikationen (z. B. auch als Ausgangsmedium für Druckausgaben) sprechen für ihre umfassende Verwendung. Und doch gibt es viele, die ihnen als wissenschaftliches Publikationsmedium skeptisch gegenüber stehen. Eine neue Kultur des elektronischen Publizierens kann sich nur durchsetzen, wenn die Langzeitarchivierung elektronischer Dokumente gesichert wird.

Ziel der Langzeitarchivierung elektronischer Dokumente muss es deshalb sein, die integrale und authentische Version der digitalen Ressourcen auch mit zukünftigen technischen Umgebungen zugänglich zu halten und benutzbar zu machen. Dabei ist mit „integer“ gemeint, dass alle signifikanten Eigenschaften des Objektes bei der Archivierungsmaßnahme erhalten geblieben sind; „authentisch“ sagt etwas über die Glaubwürdigkeit des Objektes aus: nur ein autorisiertes Objekt oder ein in einem dokumentierten Verfahren erstelltes Desiderat können als authentisch

bezeichnet werden.

Dabei bedeutet „Langzeit“ für die Bestandserhaltung digitaler Materialien – um Formulierungen von Ute Schwens und Hans Liegmann aufzugreifen – „nicht die Abgabe einer Garantierklärung über fünf oder fünfzig Jahre, sondern die verantwortliche Entwicklung von Strategien, die den beständigen, von der technischen Entwicklung verursachten Wandel bewältigen können.“

Im vorliegenden einführenden Beitrag wird wesentlich auf die beiden in Deutschland führenden Langzeitarchivierungsprojekte Nestor und Kopal eingegangen. Der Beitrag wertet dabei Materialien aus, die insbesondere von Heike Neuroth, Ute Schwens und Hans Liegmann und anderen an den Projekten unmittelbar Beteiligten erstellt wurden. Für die Erlaubnis sie zu benutzen danke ich allen.

Will man zu einer Strategie der Langzeitarchivierung kommen, muss man sich einigen, wer, wann, wofür und wie lange die Verantwortung für die dauerhafte Bereitstellung übernehmen soll. Gefragt sind dabei zunächst in erster Linie die Pflichtexemplarbibliotheken und Archive, die in der Regel einen gesetzlichen Auftrag für die langfristige Aufbewahrung von Materialien haben. Daneben sind die Sammelschwerpunktbibliotheken (Sondersammelgebietsbibliotheken und Zentrale Fachbibliotheken) zu nennen, die der Aufgabe verpflichtet sind, wissenschaftliche Literatur dauerhaft zur Verfügung zu stellen. Doch in einer Zeit, in der digitale Dokumente dezentral an vielen Stellen entstehen, sind z.B. auch die Universitäten und Forschungsinstitute aufgerufen, sich an der Aufgabe der Langzeitarchivierung zu beteiligen. Letztlich sollte aber jeder Produzent von digitalen Dokumenten Aspekte des dauerhaften Zugangs berücksichtigen.

Einigkeit herrscht im deutschen Projekt nestor, dass man nur gemeinsam die anstehenden Aufgaben bewältigen kann. Dabei wird eine Redundanz der Speicherung für unerlässlich gehalten. Kooperationen sind nicht nur zwischen Einrichtungen auf Länder- und Bundesebene notwendig; bei denen die politischen, rechtlichen und finanziellen Rahmenbedingungen für die Langzeitarchivierung geschaffen werden müssen. Auch eine interaktive Zusammenarbeit ist unumgänglich.

Ähnliche Aktivitäten wie nestor in Deutschland sind auch von anderen Ländern im Gange. Mit diesen arbeitet Nestor eng zusammen. Genannt seien:

- Digital Preservation Coalition (DPC) in Großbritannien  
<http://www.dpconline.org/>
- National Digital Information Infrastructure and Preservation Programme (NDIIP) in den USA, <http://www.digitalpreservation.gov>
- Pandora-Archiv (Australien),  
<http://pandora.nla.gov.au/>

Auf internationaler Ebene sind besonders wichtig:

- PREMIS-Arbeitsgruppe,  
<http://www.oclc.org/research/projects/pmwg/>
- International Internet Preservation Consortium (IIPC)  
<http://netpreserve.org/about/index.php>

## nestor

Im Projekt nestor (Network of Expertise in long-terms STOrage of digital Resources ) wird deshalb besonderer Wert darauf gelegt, eine breite Bewusstseinsbildung für die Problematik der Langzeitarchivierung digitaler Ressourcen zu erreichen. Das vom Bundesministerium für Bildung und Forschung (BMBF) finanzierte Projekt hat eine Laufzeit von drei Jahren (2003 – 2006). Partner sind Die Deutsche Bibliothek (DDB) als Projektleiter, die Bayerische Staatsbibliothek München (BSB), die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB Göttingen) sowie der Computer- und Medienservice der Humboldt Universität Berlin. Die Archive sind durch die Generaldirektion der Staatlichen Archive Bayern, München und die Museen durch das Institut für Museumskunde Stiftung Preußischer Kulturbesitz Berlin vertreten.

Ziele des Kompetenznetzwerkes, das diese Einrichtungen bilden, sind, ein webbasiertes deutschsprachiges Informationsforum aufzubauen. Es

sind zu entwickeln:

- Kriterien für vertrauenswürdige Archive
- Empfehlungen für die Zertifizierung digitaler Archivierung
- Empfehlung für Sammelrichtlinien und Auswahlkriterien
- Grundsätze für die Langzeitarchivierung und
- ein Konzept für die dauerhafte Organisation der Langzeitarchivierung.

Dabei ist man sich von vornherein bewusst, dass nur ein verteiltes Vorgehen Erfolg verspricht, weil keine der beteiligten Institutionen eine umfassende Archivierung leisten kann. Erforderlich aber ist ein abgestimmtes Vorgehen der Bereiche Archiv, Bibliotheks- und Museumswesen.

Die Website von nestor dient als Informations- und Kommunikationsplattform. Von besonderem Interesse für Außenstehende ist die Informationsdatenbank, die einen leichten Zugang zu relevanten Publikationen aber auch Institutionen oder Konferenzen zur Langzeitarchivierung bietet.

Nestor hat inzwischen folgende Materialien veröffentlicht:

*Digitale Langzeitarchivierung und Recht / vorgelegt von Rechtsanwältin Goebel und Scheller (Bad Homburg v.d.H.). – Frankfurt am Main : nestor c/o Die Deutsche Bibliothek, 2004. IV. 88 S.*

*Digitalisierung und Erhalt von Digitalisaten in deutschen Museen / Dirk Wittbaut. Unter Mitarbeit von Andrea Zierer, Arno Dettmers, Stefan Rohde-Enslin. – Frankfurt am Main : nestor c/o Die Deutsche Bibliothek, 2005. IV, 116 S. : graph. Darst.*

*Vergleich bestehender Archivierungssysteme / Uwe M. Borghoff u. Mitarb. Univ. d. Bundeswehr München, Fak. f. Informatik, Inst. f. Softwaretechnologie. – Frankfurt am Main: nestor c/o Die Deutsche Bibliothek, 2005. – 140 S. : graph. Darst.*

## KOPAL

Der praktischen Seite der Langzeitarchivierung dient das Forschungs- und Innovationsprojekt KOPAL, das ebenfalls vom BMBF gefördert wird. Die Laufzeit geht auch hier über drei Jahre (2004 – 2007). Ziel ist der Aufbau einer nachnutzbaren technischen und organisatorischen Infrastruktur zur Sicherung der Langzeitverfügbarkeit elektronischer Publikationen. Projektpartner sind Die Deutsche Bibliothek (DDB), die auch hier die Projektleitung innehat, die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB Göttingen), die Gesellschaft für Wissenschaftliche Datenverarbeitung Göttingen (GWDG) sowie die IBM Deutschland GmbH.

Die beiliegenden Schaubilder skizzieren die Projektpartner in Motivation, Kompetenzen und Aufgaben (Abb. 1 – 4).

**Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen** **kopal** Daten für die Zukunft

**Projektziel / Motivation**

- Langfristige Datenarchivierung ist eine Kernaufgabe der GWDG
- Langzeitarchivierung als künftig wichtige Dienstleistung der GWDG
- Großes Interesse an LZA auf Seiten beider Gesellschafter

**Kompetenzen**

- Rechen- und Kompetenzzentrum für Universität Göttingen und Max-Planck-Gesellschaft
- Backup- und Archivsystem mit räumlich getrennten Standorten der eingesetzten Bandbibliotheken
- Betreiber Hochgeschwindigkeitsnetz GÖNET (1 Gbit/sec) mit 622 Mbit/sec Anbindung an das Gigabit-Wissenschaftsnetz (G-WiN)

**Aufgaben**

- Hardware: Beschaffung + Pflege
- IT-Infrastruktur
- Serveradministration
- Speichermedienverwaltung

**Projektleitung GWDG**

- Dagmar Ullrich, dullric@gwdg.de

Die Deutsche Bibliothek | | | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN | | | IBM 37

Abb. 1: Die Rolle der Deutschen Bibliothek in kopal





**Projektziel / Motivation**

- Einbringen von Expertise und Technologien für den Aufbau einer standard-basierten Langzeitarchivierungslösung für digitale Medien. Arbeiten mit innovativen Kunden, um die Lösung eng an den fachlichen Anforderungen auszurichten.

**Kompetenzen**

- spezifische Expertise durch Aufbau von DIAS bei der Nationalbibliothek der Niederlande
- Technologien für Langzeitarchivierung als Produktschwerpunkt (z.B. Universal Virtual Computer als Basis für Emulations- / Migrationsstrategien von elektronischen Publikationen)
- Service- und Projektkompetenz

**Aufgaben**

- Lieferung von DIAS-Core und IBM Content Manager als Basis zur Langzeitarchivierung
- Funktionale und technologische Weiterentwicklung von DIAS
- Organisation und Koordinierung der europäischen DIAS User Group

**Projektleitung IBM**

- Irmgard Kurth, [ikurth@de.ibm.com](mailto:ikurth@de.ibm.com)

Die Deutsche Bibliothek      38

Abb. 4: Die Rolle der IBM Deutschland GmbH in kopalm

Softwaregrundlage ist das von der Königlichen Bibliothek in Den Haag zusammen mit IBM entwickelte DIAS (Digital Information Archiving System). Theoretische Basis von DIAS bildet das OAIS-Modell (Reference Model for an „Open Archival Information System“ – ISO 24721) (vgl. Abb. 1). Im Rahmen des EU-Projektes NEDLIB (Networked European Deposit Library) wurde die DIAS-Umsetzung durchgeführt.

SIP = Submission Information Package

AIP = Archival Information Package

DIP = Dissemination Information Package



## Lösungskonzept DIAS 1



- Theoretische Basis:
  - OAIS Modell (Reference Model for an „Open Archival Information System“ – ISO 14721)

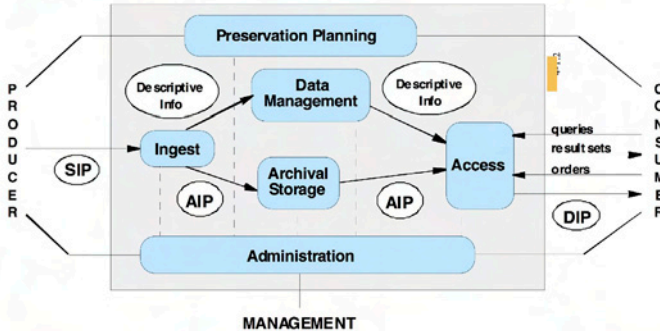


Abb. 5: Das OAIS-Modell (Reference Model for an „Open Archival Information System – ISO 24721“)

Gegenüber dem Ausgangsprodukt DIAS strebt KOPAL an, einen kooperativen Betrieb zu ermöglichen. Als neue Features werden Remote Access und Mandantenfähigkeit eingeführt. Objektformat und Datenmodell werden erweitert. Für festgelegte Gruppen von digitalen Objekten werden Werkzeuge entwickelt und implementiert. Ziel ist dabei, die Einlieferungsroutinen zu automatisieren und sie für große Dokumentmengen zu optimieren.

## EMANI

Die Electronic Mathematic Archives Network Initiative (EMANI) strebt an, für mathematische Publikationen trusted repositories aufzubauen. In internationaler Kooperation soll die Langzeitverfügbarkeit und der Zugriff auf digitale wissenschaftliche Publikationen in der Mathematik gesichert werden. Initiativeeinrichtungen sind unter der Federführung des Zentralblatts der Mathematik

- die Bibliothek der Cornell University,
- die Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB Göttingen),
- die Tsinghua University Library (Beijing) und
- die Orsay Mathematical Library.

Inhalte stellen derzeit bereit:

- Springer Verlag
- Birkhauser-Verlag
- Teubner-Verlag
- Vieweg-Verlag sowie
- ELib M in EMIS

Bei der Langzeitarchivierung ist die Definition geeigneter Metadaten von besonderer Bedeutung. Hier seien als wichtige Teilaspekte genannt:

- Strukturelle Metadaten beschreiben Relationen zu anderen Objekten im Archiv
- Administrative Metadaten beschreiben den Lebenszyklus des Objekts (z.B. vorgenommene Archivierungsmaßnahmen)
- Rechtliche Metadaten beschreiben Zugriffsmodalitäten auf das Objekt
- Technische Metadaten beschreiben die erforderliche technische Umgebung.

Auch hier besteht nationale und internationale Zusammenarbeit mit dem Ziel, zu einheitlichen Lösungen zu kommen .

Bei der Langzeitarchivierung werden zwei Strategien unterschieden, die Migration und die Emulation.

Die Migration beruht auf der Transformation eines digitalen Objekts von einer technischen Umgebung in eine andere. Vorteil des Verfahrens ist die relativ etablierte Prozedur.

Nachteile sind, dass die Migration

- oft nicht verlustfrei möglich ist,
- ein unendlicher Prozess ist, dessen Kosten schwer kalkulierbar sind. Außerdem ist er
- für komplexe digitale Objekte ungeeignet.

Die Emulation strebt die Nachbildung der für die Benutzung erforderlichen originalen Umgebung an.

Vorteile sind:

- Das digitale Objekt bleibt unverändert erhalten
- Spezifikationen der technischen Umgebung werden mitarchiviert
- Die Authentizität des digitalen Objekts ist gewährleistet

Nachteil ist, dass das Verfahren sehr aufwändig ist und bisher nur in sehr begrenztem Ausmaß durchgeführt werden kann. Im Kopal Projekt sind Ansätze des Einsatzes der Emulation gegeben.

Zusammenfassend lässt sich sagen, dass in Deutschland mit dem Projekt Nestor in internationaler Zusammenarbeit die Methoden der Langzeitarchivierung entwickelt und ihre organisatorische Durchführung vorbereitet wird. Mit dem Projekt Kopal ist ein zukunftsweisendes Angebot für die praktische Umsetzung in Entwicklung, die von vielen Interessierten wird angewendet werden können.