

nestor Handbuch:  
**Eine kleine Enzyklopädie  
der digitalen Langzeitarchivierung**  
6 Organisation

## Herausgeber

Heike Neuroth  
Hans Liegmann †  
Achim Oßwald  
Regine Scheffel  
Mathias Jehn  
Stefan Strathmann

GEFÖRDERT VOM



Bundesministerium  
für Bildung  
und Forschung

## Im Auftrag von

nestor – Kompetenznetzwerk Langzeitarchivierung und Langzeitverfügbarkeit  
digitaler Ressourcen für Deutschland  
nestor – Network of Expertise in Long-Term Storage of Digital Resources  
<http://www.langzeitarchivierung.de>

## Kontakt

[editors@langzeitarchivierung.de](mailto:editors@langzeitarchivierung.de)

c/o

Niedersächsische Staats- und Universitätsbibliothek Göttingen

Dr. Heike Neuroth

Forschung und Entwicklung

Papendiek 14

37073 Göttingen

Tel. +49 (0) 55 1 39 38 66

Der Inhalt steht unter folgender Creative Commons Lizenz:  
<http://creativecommons.org/licenses/by-nc-sa/2.0/de/>



## 6 Organisation

*Christian Keitel*

### **Einführung**

Die Organisation der digitalen Langzeitarchivierung kann als die arbeitsteilige Bewältigung dieser Aufgabe verstanden werden. Die bisherigen Erfahrungen zeigen, dass die bei der digitalen Langzeitarchivierung anstehenden (Teil-)Aufgaben sehr unterschiedlich organisiert und voneinander abgegrenzt werden können. Zumeist erfolgt eine Aufgabenteilung zwischen einzelnen, mit der Langzeitarchivierung befassten Institutionen. Modelle zur Arbeitsteilung innerhalb einer Institution (interne Organisationsmodelle) werden kaum veröffentlicht, zumal die Aktivitäten derzeit oft nur einen zeitlich befristeten Projektstatus besitzen. Eine allgemeine und generalisierende Darstellung der Organisation ist daher nur auf einem sehr abstrakten Niveau möglich. Einen solchen Rahmen bietet das Funktionsmodell (*functional modell*) von OAIS. Anschließend an diese aufgabenorientierte Beschreibung werden weitere Faktoren genannt, die bei der Organisation der digitalen Langzeitarchivierung zu berücksichtigen sind. In einem dritten Schritt werden anhand konkreter Beispiele mögliche Umsetzungen skizziert.

## 1. Aufgaben nach dem OAIS-Funktionsmodell

**Produktion:** Die Entstehung der Daten ist nach OAIS nicht Bestandteil eines digitalen Langzeitarchivs. Aus diesem Grund müssen im Bereich *Ingest* Schnittstellen und Übergabe- bzw. Aufnahme-prozeduren detailliert beschrieben werden. In den meisten Fällen ist das digitale Archiv organisatorisch vom Produzenten getrennt. Manchmal wird diese Trennung aber auch relativiert oder aufgehoben:

Archivierung durch die Produzenten (1): 1996 wurde den australischen Behörden nach der Theorie des *records continuum* auferlegt, alle alten, im Dienst nicht mehr benötigten, Dokumente dauerhaft selbst zu verwahren. Den Archiven kam dabei die Rolle zu, das Funktionieren des Konzepts sicherzustellen, also eine Art „Archivierungspolizei“ zu spielen. Bereits 2000 kehrte das Australische Nationalarchiv wieder zu seiner traditionellen Politik zurück, d.h. zur Übernahme dieser Dokumente. Nur Archive und Bibliotheken haben ein genuines Interesse an der Erhaltung von Informationen, die in den Augen ihrer Ersteller „veraltet“ sind. Erst dieses Interesse gewährleistet, dass vermeintlich uninteressante Daten weiterhin gepflegt werden.

Archivierung durch die Produzenten (2): Die Systeme der Umweltbeobachtung verwahren aktuell produzierte Daten zusammen mit den Daten vergangener Jahrzehnte. Die einzelnen Informationen sollen dauerhaft im selben System und unter denselben Namen aufgefunden und angesprochen werden, die systemische Einheit dieser Daten ist über einen langen Zeitraum hinweg erwünscht. Die Information veraltet also im Gegensatz zum beschriebenen australischen Beispiel theoretisch nie. Vergleichbare Systeme werden derzeit in vielen Naturwissenschaften aufgebaut. Ist es nicht ganz allgemein sinnvoll, bei der Entstehung der Daten dieselben Erhaltungsregeln anzuwenden wie später im Archiv? Analog hierzu sieht sich das 2003 gegründete britische *Digital Curation Centre* auch für den ganzen Lifecycle eines Dokuments zuständig: “The term ‘digital curation’ is increasingly being used for the actions needed to maintain and utilise digital data and research results over their entire life-cycle for current and future generations of users.”<sup>1</sup>

Archivisches Engagement bei den Produzenten: Seit über 15 Jahren engagieren sich die klassischen Archive in den Behörden bei der Einführung elektronischer Akten und anderer digitaler Systeme. Ihr Motiv: Bei der Einführung eines Systems werden die Grundlagen dessen gelegt, was

---

1 JISC Circular 6/03 (Revised), in: <http://www.dcc.ac.uk/docs/6-03Circular.pdf>.

dann später im Archiv ankommt. Danach ist es weniger aufwändig, in der Behörde Dinge grundsätzlich zu regeln, als später jedes Objekt einzeln nachbearbeiten zu müssen. Im DOMEA-Konzept (**D**okumentent**m**anagement und **E**lektronische **A**rchivierung) werden die beiden Bereiche auch begrifflich zusammengezogen.

Archive werden zu Produzenten: Durch die Digitalisierungsstrategien der Archive und Bibliotheken mutieren diese klassischen Gedächtnisinstitutionen auf einmal selbst zu Datenproduzenten. Es bedarf zwar zusätzlicher Qualitätssicherungsmaßnahmen für die Digitalisate, eine Ingest-Schnittstelle oder die Umwandlung von SIPs in AIPs sind jedoch nicht mehr notwendig.

**Ingest**: Setzt man mit OAIS eine Trennung zwischen Produktions- und Archivzuständigkeit, dann müssen im Ingest die Übernahmepakete (SIPs) entgegen genommen, überprüft, und in Archivierungspakete (AIPs) umgewandelt werden. Beschreibende Metadaten werden extrahiert und an das Data Management weitergegeben. Der ebenfalls von den OAIS-Autoren verfasste PAIMAS-Standard gliedert diesen Bereich in insgesamt vier Phasen: Nach einer Vorbereitungsphase werden in einer Definitionsphase alle wesentlichen Rahmenbedingungen vereinbart und erprobt. Hierzu gehört insbesondere die Auswahl der zu übernehmenden Objekte und die Abklärung sämtlicher rechtlichen Aspekte. Während der Transferphase werden diese übernommen und schließlich in der Validierungsphase auf ihre angenommenen Eigenschaften hin überprüft.

Auch bei einer festen Trennung zwischen Produzenten und Archiv können die einzelnen Aufgaben sehr unterschiedlich aufgeteilt werden. Hierzu gehören die Auswahl der Objekte, ihre Ausstattung mit Metadaten und die ggf. erforderliche Migration der Objekte in ein archivierungsfähiges Format. Entsprechend kann sich die dem Archiv verbleibende Ingest-Aufgabe v.a. administrativ gestalten (d.h. es gibt dem Produzenten die entsprechenden Vorgaben) oder zunehmend auch technische Komponenten enthalten (d.h. es setzt diese Punkte selbst um). Die Entscheidung für eine der beiden Optionen ist wesentlich von der Gleichartigkeit der Objekte abhängig: Erst wenn sich die Objekte sehr stark gleichen, kann die Zahl der Vorgaben so weit reduziert werden, dass eine entsprechende Automatisierung auch erfolgreich umgesetzt werden kann. Bei stark differierenden Objekten lassen sich diese Regeln nicht in einer vergleichbar umfassenden Weise aufstellen, weshalb die Aufgaben vom Archiv selbst übernommen werden müssen, was dessen Aufwand entsprechend erhöht.

Im letztgenannten Fall können dann weitere Teilaufgaben gebildet werden. Bei-

spielsweise kann die Metadatenerfassung in zwei aufeinanderfolgende Schritte aufgespalten werden: a) Anlegen erster identifizierender Metadaten. b) Nähere Beschreibung im Zuge der weiteren Bearbeitung.

**Archival Storage:** In diesem Bereich werden die AIPs über einen langen Zeitraum gespeichert. Der Zustand der Speichermedien wird kontinuierlich überwacht, im Bedarfsfall werden einzelne Medien ersetzt, regelmäßig werden auch ganze Medien-Generationen in neuere Speichertechnologien migriert. Neben Hardware und Software sind hier also v.a. IT-Kenntnisse erforderlich. Es ist daher auch der Bereich, der am ehesten von den klassischen Gedächtnisinstitutionen an externe Rechenzentren ausgelagert wird. Andererseits unterscheiden sich die Anforderungen der digitalen Langzeitarchivierung z.T. erheblich von denen, die gewöhnlich an Rechenzentren gestellt werden. Die National Archives and Records Administration (NARA) der Vereinigten Staaten hat daher Anfang der 1990er Jahre den Bereich wieder ins Haus geholt.

**Data Management:** In diesem Bereich werden die identifizierenden, beschreibenden und administrativen Metadaten gepflegt. Er ist daher für die klassischen Gedächtnisinstitutionen nicht neu. Sofern nicht ein eigenes Recherchesystem für die digitalen Objekte aufgebaut wird, liegt es nahe, dass dieser Bereich von den Organisationseinheiten übernommen wird, die bereits für die Beschreibung der analogen Objekte zuständig sind.

**Preservation Planning:** Digitale Langzeitarchivierung erfordert eine kontinuierliche aktive Begleitung der archivierten Objekte. Zentral ist die Terminierung und Koordination der einzelnen Erhaltungsprozesse. Schnittstellen bestehen zu den Bereichen Ingest, Archival Storage und Data Management.

**Access:** Diese Einheit ermöglicht die Benutzung der digitalen Objekte. Sie ermöglicht die Recherche in den beschreibenden Metadaten und liefert die Benutzungspakete aus (DIPs). Manche Archive überlassen diese Aufgabe aber auch ihren Benutzern, d.h. ausgegeben werden die nicht weiter veränderten AIPs.

**Administration:** Der Bereich klärt das Zusammenspiel der einzelnen Organisationseinheiten. Er handelt grundsätzliche Vereinbarungen mit den Produzenten aus, definiert die Rahmenbedingungen für eine Benutzung, überwacht das Archivsystem, entwickelt Standards und Policies und berichtet regelmäßig dem außerhalb des OAI angesiedelten Management. Er ist somit kaum technisch geprägt.

## 2. Weitere Faktoren

Die Organisation der digitalen Langzeitarchivierung ist außer von den Aufgaben und den zu archivierenden Objekten auch von weiteren Faktoren abhängig. Genannt werden können die Größe der Einrichtung, ihre sonstigen Aufgaben und die Qualifikation ihres Personals. Sehr große Archive können zu jeder Einheit des OAIS-Funktionsmodells mindestens eine administrative Einheit bilden. Zusätzlich kann noch ein Forschungsbereich ausgegliedert werden. Kleinere Archive sind dagegen gezwungen, mit weniger administrativen Einheiten auszukommen. Bei klassischen Gedächtniseinrichtungen stellt sich die Frage, welche Aufgaben unabhängig vom Medientyp bearbeitet werden können. In zahlreichen Bereichen sind zudem sowohl die Kenntnisse traditionell ausgebildeter Archivare oder Bibliothekare als auch ausgeprägte IT-Kenntnisse erforderlich. Die Organisation ist daher auch von dem bereits bestehenden Personalbestand der Einrichtung und der Möglichkeit einer Neueinstellung abhängig.

## 3. Beispiele/Umsetzung in die Praxis

### 3.1. Centre national d'études spatiales (CNES)

Die französische Raumfahrtagentur CNES archiviert fast ausschließlich digitale Daten. Es wurden drei administrative Einheiten gebildet: a) Ingest, b) Archival Storage und c) Data Management und Access. Im Ingest arbeiten Archivare und Computerspezialisten zusammen. Der Archivar definiert die zu übernehmenden Objekte, überprüft sie auf ihre Vollständigkeit und strukturiert sie. Der Computerspezialist definiert Daten und Metadaten, nimmt die physische Übernahme und die Validierung vor und entwickelt entsprechende Tools. Das neue Berufsbild des *Digital Data Manager* kann auf beiden Gebieten des Ingest tätig werden. Beim Archival Storage werden ausschließlich Computerspezialisten eingesetzt. Seit 1994 wird dieser Bereich vom STAF (*Service de Transfert et d'Archivage de Fichiers*) ausgeführt. Die OAIS-Bereiche Data Management und Access werden beim CNES zusammengezogen. Im Vordergrund stehen Datenbank-, Retrieval- und Internettechnologien, daneben werden vertiefte Kenntnisse über das Archiv benötigt. Das Funktionieren des Archivs wird durch eine Koordinationsstelle, bewusst klein konzipierte Überlappungsbereiche und die weitgehende Unabhängigkeit der einzelnen Einheiten gewährleistet.

### 3.2 The National Archives (UK)

Die National Archives haben mehrere objektspezifische Ansätze zur digitalen Archivierung entwickelt, die zusätzlich von zentralen Systemen (z.B. die Formatdatenbank PRONOM) unterstützt werden. Seit 2001 ist zudem für die Erhaltung von *born digital material* nicht mehr das *Records Management Department* sondern das neu eingerichtete *Digital Preservation Department* zuständig. Für strukturierte Daten wurde 1997 eine Kooperationsvereinbarung mit dem Rechenzentrum der Londoner Universität (University of London Computer Centre, UCLL) geschlossen, in deren Folge das *National Digital Archive of Datasets* (NDAD) 1998 in Betrieb genommen werden konnte. Die National Archives sind für die Auswahl der Daten und die Definition der Service-Levels zuständig, NDAD für alle weiteren Aufgaben (explizit unterschieden werden Ingest, Preservation und Access). Im NDAD arbeiten zwölf Personen in vier Disziplinen: Die *Project Archivists* treffen zentrale Entscheidungen über die Organisation des Archivs, Katalogisierung und Indexierung und leiten die Computer-Spezialisten an. Die *Archive Assistants* sind für die Benutzerbetreuung zuständig. Sie unterstützen die Project Archivists z.B. durch Einscannen der Papierdokumentation. Die *Data Specialists* sind für die technische Umsetzung der getroffenen Entscheidungen zuständig, Der *Systems Support Staff* stellt schließlich das Funktionieren von Hard- und Software sicher. Für die Archivierung elektronischer Records (Akten) wurde in den National Archives Mitte der 1990er Jahre das EROS-Projekt aufgesetzt, das nun im Seamless-Flow-Programm fortgesetzt wird. Erste Ergebnisse sind ab Ende 2007 zu erwarten. Gleichzeitig werden im 2003 in den National Archives gegründeten *Digital Archive* bereits Records übernommen und Erfahrungen aufgebaut. Für die Archivierung von Internetseiten haben sich die National Archives 2003 mit der British Library, den Nationalbibliotheken von Wales und Schottland, JISC und dem Wellcome Trust zum *UK Web Archiving Consortium* zusammengeschlossen, um eine gemeinsame Infrastruktur zur Web-Archivierung aufzubauen<sup>2</sup>.

### 3.3 Deutsche Nationalbibliothek (DNB) und Staats- und Universitätsbibliothek Göttingen (SUB)

Die Deutsche Nationalbibliothek und die Staats- und Universitätsbibliothek Göttingen haben ihre Lösung zur Archivierung digitaler Objekte im Projekt

---

2 Auf einer vergleichbaren Kooperation basiert das BOA-Projekt. Die beiden Landesbibliotheken und das Landesarchiv von Baden-Württemberg sind zuständig für die Auswahl und den Ingest der Webseiten und Einzel-Dokumente, während das Bibliotheksservicezentrum Baden-Württemberg das Speichersystem und die Infrastruktur zur Verfügung stellt.



KOPAL gemeinsam mit der Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) und der IBM Deutschland entwickelt. Die Partner gehen von einem arbeitsteiligen Vorgehen aus: Die Übernahme und Aufbereitung der AIPs liegt in den Händen der beteiligten Bibliotheken und erfolgt durch eine OpenSource-Software. Die fertigen AIPs werden dann per Fernzugriff zentral im Rechenzentrum der GWDG gespeichert. Dabei kommt das durch die IBM entwickelte DIAS-System zu Einsatz. Die Benutzung erfolgt dann wiederum durch Fernzugriff bei den beiden Bibliotheken.

## Literatur

- Reference Model for an Open Archival Information System (OAIS), Blue Book 2002, <http://www.ccsds.org/publications/archive/650x0b1.pdf>
- Producer-Archive Interface Methodology Abstract Standard, Blue Book 2004, <http://public.ccsds.org/publications/archive/651x0b1.pdf> (PAIMAS)
- DOMEA-Konzept: Das Organisationskonzept, die Erweiterungsmodule und weitere Informationen finden sich auf den Seiten der KBSt: <http://www.kbst.bund.de>.
- Adrian Brown, Developing practical approaches to active preservation, in: Proceedings of the 2nd International Conference on Digital Curation, Glasgow 2006
- Adrian Brown, Archiving Websites. A Practical Guide for Information Management Professionals, London 2006
- Claude Huc, An organisational model for digital archive centres, [http://www.erpanet.org/events/2004/amsterdam/presentations/erpaTraining-Amsterdam\\_Huc.pdf](http://www.erpanet.org/events/2004/amsterdam/presentations/erpaTraining-Amsterdam_Huc.pdf); auch als ...ppt und ...m3u.
- Richard Jones, Theo Andrew, John MacColl, The Institutional Repository, Oxford 2006
- Patricia Sleeman, It's Public Knowledge: The National Digital Archive of Datasets, in: Archivaria 58 (2004), S. 173 - 200

