

H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth (Hrsg.)

# nestor Handbuch

Eine kleine Enzyklopädie  
der digitalen Langzeitarchivierung

Version 2.3

Kapitel 18  
Praxisbeispiele

nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung  
hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, K. Huth  
im Rahmen des Projektes: 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

Bibliografische Information der Deutschen Nationalbibliothek  
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen  
Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter  
<http://www.d-nb.de/> abrufbar.

Neben der Online Version 2.3 ist eine Printversion 2.0 beim Verlag Werner Hülsbusch,  
Boizenburg erschienen.

Die digitale Version 2.3 steht unter folgender Creative-Commons-Lizenz:  
„Namensnennung-Keine kommerzielle Nutzung-Weitergabe unter gleichen Bedingungen 3.0  
Deutschland“  
<http://creativecommons.org/licenses/by-nc-sa/3.0/de/>



Markenerklärung: Die in diesem Werk wiedergegebenen Gebrauchsnamen, Handelsnamen,  
Warenzeichen usw. können auch ohne besondere Kennzeichnung geschützte Marken sein und  
als solche den gesetzlichen Bestimmungen unterliegen.

URL für Kapitel 18 „Praxisbeispiele“ (Version 2.3): [urn:nbn:de:0008-20100624173](http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-20100624173)  
<http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-20100624173>



*Gewidmet der Erinnerung an Hans Liegmann (†), der als Mitinitiator und früherer Herausgeber des Handbuchs ganz wesentlich an dessen Entstehung beteiligt war.*

## 18 Praxisbeispiele

### 18.1 Einführung

*Regine Scheffel*

nestor hat als Kompetenzzentrum in Sachen Langzeitarchivierung digitaler Objekte ganze Arbeit geleistet: Fachleute aus dem Kulturerbebereich haben das neue Arbeitsfeld angenommen, Projekte und erste Erfahrungsberichte zeigen, dass auf unterschiedlichen Gebieten Lösungen für die Herausforderung, unser digitales Erbe zu erhalten, erarbeitet werden. Neben Archiven, Bibliotheken und Museen arbeiten wissenschaftliche Einrichtungen, Firmen, Organisationen und Behörden daran die digitale Langzeitarchivierung in Organisationsstruktur, Aufgabenbereichen und Arbeitsprozessen zu verankern. Noch sind viele auf dem Weg und betrachten ihre Lösungen nicht als endgültig abgeschlossen. Das ist bei diesem Arbeitsfeld wohl auch gar nicht anders möglich, zu stark sind die zu erhaltenden Objekte, aber auch Konzepte und technische Lösungen zur digitalen Langzeitarchivierung Entwicklungsprozessen unterworfen.

Dennoch freut sich das Herausgeberteam die ersten Praxisbeispiele vorstellen zu können:

Die Deutsche Nationalbibliothek (DNB) hat in Deutschland die Vorreiterrolle gespielt bei der Einrichtung eines Archivsystems. Mit kopal liegt nun ein System vor, das nicht nur die große Menge digitaler Publikationen für die DNB archiviert, sondern auch anderen Einrichtungen offen steht.

Das Bibliotheksservicezentrum (BSZ) in Konstanz hat die Bildarchivierung für die Staatsgalerie Stuttgart auf Dauer übernommen. Noch ist das Archivsystem nicht abschließend realisiert, doch sind wesentliche OAIS-konforme Festlegungen und Workflows definiert und stehen weiteren Einrichtungen zur Nachnutzung zur Verfügung.

Der Blick über den Tellerrand zeigt, dass auch außerhalb von Landes- und Bundeseinrichtungen der Funke überggesprungen ist: Der Bundestag hat sich damit begonnen seine Internetangebote in transparenter Auswahl in ein langzeitarchivierungstaugliches Webarchiv zu überführen und arbeitet nun an der Optimierung des Systems.

Man darf gespannt darauf sein, welche Einrichtungen oder Firmen ihre Praxis der digitalen Langzeitarchivierung in den nächsten Ausgaben dieses nestor Handbuchs vorstellen werden.

## 18.2 Langzeitarchivierung von elektronischen Publikationen durch die Deutsche Nationalbibliothek

Maren Brodersen und Sabine Schrimpf <sup>1</sup>

*Seit das Gesetz über die Deutsche Nationalbibliothek vom 22. Juni 2006 in Kraft getreten ist, erstreckt sich der Sammelauftrag der Deutschen Nationalbibliothek auch auf „Medienwerke in unkörperlicher Form“, d.h. auf Netzpublikationen. Konkret hat sie den Auftrag, alle in Deutschland veröffentlichten und deutschsprachigen Medienwerke „zu sammeln, zu inventarisieren, zu erschließen und bibliografisch zu verzeichnen, auf Dauer zu sichern und für die Allgemeinheit nutzbar zu machen“.<sup>2</sup> Dabei gelten als Medienwerke alle Darstellungen in Schrift, Bild und Ton, die in körperlicher Form verbreitet werden (d.h. auf Papier, elektronischen oder anderen Datenträgern) oder in unkörperlicher Form über öffentliche Netze, in der Regel das Internet, zugänglich gemacht werden.*

*Für diese Situation werden Verfahren für die Sammlung und Langzeitarchivierung von Netzpublikationen an der Deutschen Nationalbibliothek permanent weiter entwickelt und implementiert. Dieser Artikel kann daher lediglich einen Überblick über den aktuellen Stand der Sammlung und Langzeitarchivierung von Netzpublikationen an der Deutschen Nationalbibliothek geben. Es wird eingegangen auf den in Pflichtablieferungsverordnung und Sammelrichtlinien näher bestimmten Sammelauftrag der Deutschen Nationalbibliothek, auf speziell für Netzpublikationen entwickelte Ablieferungs- und Erschließungsverfahren und die Langzeitarchivierung von Netzpublikationen.*

### Sammelgebiet

Nicht jede deutschsprachige Webseite im Internet gehört automatisch zum Sammelauftrag der Deutschen Nationalbibliothek. Zwei Dokumente regeln die Einzelheiten zum Sammelgebiet Netzpublikationen und schränken die Ablieferungspflicht nach bestimmten Selektionskriterien ein: die Pflichtablieferungsverordnung und die Sammelrichtlinien.

---

1 Mit Unterstützung von Sarah Hartmann, Susanne Puls und Tobias Steinke.

2 Gesetz über die Deutsche Nationalbibliothek (DNBG) vom 22. Juni 2006, hier § 2 Abs. 1, veröffentlicht im Bundesgesetzblatt Jahrgang 2006 Teil I Nr. 29, ausgegeben zu Bonn am 28. Juni 2006, verfügbar unter <http://bundesrecht.juris.de/dnbg/index.html>

In der Pflichtablieferungsverordnung vom 17. Oktober 2008 (PflAV)<sup>3</sup> werden eine Reihe von Netzpublikationen von der Ablieferungspflicht ausgenommen, darunter Publikationen, die nicht von besonderem öffentlichen Interesse sind, wie Netzpublikationen, die lediglich privaten oder gewerblichen Zwecken dienen, Netzpublikationen, die lediglich einer privaten Nutzergruppe zugänglich sind oder Netzpublikationen von Kreisen, Gemeinden und Gemeindeverbänden, die ausschließlich amtlichen Inhalt enthalten. Auch Vorabveröffentlichungen, reine Software- oder Anwendungstools, Fernseh- und Hörfunkproduktionen und Spiele fallen nicht unter den Sammelauftrag der Deutschen Nationalbibliothek. Ebenfalls nicht sammelpflichtig sind E-Mail-Newsletter, sofern sie kein Webarchiv haben und Kommunikations-, Diskussions- oder Informationsinstrumente ohne sachliche oder personenbezogene Zusammenhänge.

Ist ein Werk parallel als Printausgabe und Netzpublikation erschienen, so sind sowohl das gedruckte Werk als auch die Netzpublikation sammelpflichtig und müssen an die Deutsche Nationalbibliothek abgeliefert bzw. von ihr eingesammelt werden. Bei unterschiedlichen technologischen, sonst aber inhaltsgleichen Ausführungen von Netzpublikationen genügt die Ablieferung bzw. Sammlung einer Version.

Näher ausgeführt werden die Auswahlkriterien in den Sammelrichtlinien. Die Sammelrichtlinien haben Handreichungscharakter für die Bibliothekare und enthalten klare Anweisungen, wie beispielsweise: „Zu sammeln sind Netzpublikationen mit Themen- oder Personenbezug, wie z.B. Netzpublikationen von und über Persönlichkeiten des öffentlichen Lebens; dazu gehören insbesondere Politiker, Schauspieler, Musiker, Schriftsteller, Maler, Wissenschaftler, Publizisten, Journalisten usw.“ Die Sammelrichtlinien waren zur Drucklegung dieses Werkes noch nicht veröffentlicht, können nach ihrer Veröffentlichung aber auf der Website der Deutschen Nationalbibliothek eingesehen werden.<sup>4</sup>

## Selektion und Praxis der Ablieferung

Neben den Sammelrichtlinien spielen die Auswirkungen auf die etablierten Geschäftsgänge eine große Rolle, denn mit der Verbreitung des digitalen Publizierens ist auch ein Wandel der bisherigen Vertriebs- und Verarbeitungswege

---

3 Verordnung über die Pflichtablieferung von Medienwerken an die Deutsche Nationalbibliothek vom 17. Oktober 2008, veröffentlicht im Bundesgesetzblatt Jahrgang 2008 Teil I Nr. 47, ausgegeben zu Bonn am 22. Oktober 2008, verfügbar unter <http://www.bgblportal.de/BGBL/bgbl1f/bgbl108s2013.pdf>.

4 Die Sammelrichtlinien können unter <http://www.d-nb.de/netzpub/index.htm> <urn:nbn:de:101-2009033003> eingesehen werden.

verbunden. Für die Deutsche Nationalbibliothek verändert sich damit die Selektion der Abnehmer von Netzpublikationen. In der Printwelt sind die Abnehmer in der Regel Verlage, wirtschaftliche und wissenschaftliche Institutionen und Organisationen sowie ein kleiner Kreis von Privatpersonen; die Vertriebsstrukturen sind seit Jahrzehnten unverändert. Die Verlage sind bekannt und ein großer Teil der traditionellen Publikationen wird über das VLB (Verzeichnis Lieferbarer Bücher)<sup>5</sup> gemeldet. Teilweise sind die Abnehmer von Netzpublikationen identisch mit den bereits bekannten Abnehmern von Printpublikationen. Das Internet erweitert jedoch den Kreis der zur Ablieferung verpflichteten Produzenten um ein vielfaches und führt in gewisser Hinsicht zu einer Anonymisierung. De facto kann jeder zum Autor und damit zum Produzent von Netzpublikationen werden.

Wie spricht die Deutsche Nationalbibliothek diesen neuen Typ von Produzenten an? Eine Möglichkeit ist die Anmeldung als Abnehmer von Netzpublikationen über ein Webformular, das auf der Website der Deutschen Nationalbibliothek bereitgestellt wird. Die Anmeldung ist offen für jedermann. Nach der Übermittlung der Adressdaten überprüfen Bibliotheksmitarbeiter die Angaben und schalten die Produzenten für die Ablieferung frei. Im Rahmen von Veröffentlichungen zum Thema Netzpublikationen, wie beispielsweise zur Ankündigung der PflAV oder in Workshops, die zum Thema organisiert werden, wird dieses Verfahren erläutert.

## Dynamische Entwicklung von Ausgabeformaten

Auch was die Ausgabeformate oder –formen betrifft, ist der traditionelle Publikationsmarkt im Umbruch. Große Verlage wie z.B. Springer arbeiten seit Jahren an der Optimierung ihrer Netzpublikationen und der entsprechenden Anpassung ihrer Geschäftsgänge. Wurde hier bis vor kurzem noch die Printpublikation zuerst auf dem Markt angeboten und erst danach die Netzpublikation über die Verlagsplattform, dann ist dies inzwischen umgekehrt der Fall.

Die dynamischen Entwicklungen des Internets und der damit verbundenen Technologien stellen große Herausforderungen für die Selektion, Sammlung und Langzeitarchivierung von Netzpublikationen dar, weil sich alle Planungen auf ein bewegtes Ziel richten. Galt beispielsweise lange Zeit das eBook<sup>6</sup> als die klassische Form der Netzpublikation, so werden die Endgeräte vielfältiger und

---

5 Informationen zum Verzeichnis Lieferbarer Bücher: <http://www.vlb.de>

6 Als eBook werden einerseits Netzpublikationen bezeichnet, die ein spezielles Lesegerät benötigen, häufig aber auch nur PDF-Dateien, die als Onlineversion die Printpublikationen abbilden.



mobiler und damit wächst die Suche nach Formaten, die multifunktional einsetzbar sind, wie beispielsweise XML oder auch das eigens für diesen Zweck entwickelte epub-Format.<sup>7</sup> Das hat für die Verlage zur Folge, dass auch die gewohnten Vertriebswege erweitert werden müssen, da sich die Zielgruppen verändern und damit die unterschiedlichsten Ansprüche und Erwartungen haben. Zunehmend wird nach Lösungen gesucht und in Form eigener Portalentwicklungen mit individueller Hard- und Software gefunden. Die Deutsche Nationalbibliothek steht vor der Herausforderung, Ablieferungsverfahren zu entwickeln, die mit diesen unterschiedlichen Portalsystemen harmonisieren.

Auffällig ist, dass sich auf anderer Ebene ein Trend zur Standardisierung abzeichnet und zwar im Bereich der Metadaten. So wird beispielsweise im Verlagswesen seit 2000 der Metadatenstandard ONIX<sup>8</sup> entwickelt, um über dieses Datenformate verschiedene Geschäftsgänge zu bedienen: Informationen auf der Website, Daten für die Meldung an das VLB, ggf. auch Daten für die Presse bzw. die Verkaufskataloge.

### **Geschäftsgänge für die Sammlung von Netzpublikationen**

Betrachtet man die Erfahrungen bei der Sammlung von Netzpublikationen auf freiwilliger Basis und die Entwicklung in diesem Bereich über die vergangenen sieben Jahre, dann wird deutlich, dass Geschäftsgänge, die einmal entwickelt wurden, um Netzpublikationen einzusammeln, steten Veränderungen unterliegen. Berücksichtigt man dann noch die unterschiedlichen Mengen, die produziert werden, dann zeigt sich auch hier, dass unterschiedliche Ablieferungsverfahren und damit Geschäftsgänge für die Verarbeitung erforderlich sind.

Aus den Erfahrungen mit den unterschiedlichen Dateiformaten und den verschiedenen Ablieferungsverfahren wurden deshalb neue Anforderungen abgeleitet und spezielle Geschäftsgänge entwickelt. Im Vordergrund stand ein pragmatischer Ansatz mit Konzentration auf das einzeln zu adressierende Objekt, d.h. die Netzpublikation mit Entsprechung in der Printwelt, die sog. druckbildähnliche Netzpublikation, die in der Regel im PDF-Format erscheint. Die nach wie vor verbreitete Trennung in Monografien und Zeitschriften ermöglicht die Orientierung an der Printwelt. Ein einheitliches Dateiformat erleichtert zudem die Ablieferung und ermöglicht automatisierte Prüfroutinen. Zentrales Ziel war in erster Linie die Automatisierung der verschiedenen Geschäftsgänge auf Bibliotheksseite: den automatisierten Import von Netzpublikationen in ein Archivsystem, die automatisierte Vergabe einer URN als Persistent Iden-

---

7 Informationen zum epub-Format: <http://www.idpf.org>

8 Informationen zu ONIX als Metadatenstandard: <http://www.editeur.org/onix.html>

tifier (wenn die Netzpublikation keinen besitzt) und den Import von Metadaten in das eigene Katalogsystem, um hier einen Datensatz zu erstellen. Für Zeitschriften bedeutet dies die Ablieferung auf Heft- oder Articlebene. Im Formular erfolgt die Verknüpfung über den in einer Auswahlliste angezeigten Zeitschriftentitel. Damit können Zeitschriften auf Heft- oder Articlebene recherchiert und angezeigt werden.

Der automatisierte Import der Netzpublikation erfolgt über eine sog. Transfer-URL. Hier kann direkt auf das PDF zugegriffen und die Netzpublikation „abgeholt“ werden. Wurde im Formular kein eindeutiger Identifier angegeben, dann wird an dieser Stelle auch eine URN der Deutschen Nationalbibliothek automatisch vergeben. Der Identifier ist das Bindeglied zwischen dem Katalogdatensatz und der Netzpublikation auf dem Archivsystem.

In einem Metadaten-Kernset<sup>9</sup> ist festgelegt, welche Metadaten erforderlich sind, um einen Kerndatensatz im Katalogsystem zu erstellen. Darüber hinaus sind weitere Metadaten festgelegt, deren Lieferung wünschenswert ist. Im Webformular sind die Kerndaten als Pflichtfelder festgelegt. Die angegebenen Metadaten können unmittelbar geprüft und ggf. korrigiert werden. Die über das Formular erfassten Metadaten werden dann in das Katalogsystem importiert und die Anzeige im Katalog erfolgt umgehend.

Eine Anzeige in der Deutschen Nationalbibliografie<sup>10</sup> erfolgt allerdings erst nach der Formal- und Sacherschließung; d.h. nach einer intellektuellen Erschließung anhand der geltenden Regelwerke und der Verknüpfung mit den Normdateien PND (Personennamendatei)<sup>11</sup> und GKD (Gemeinsame Körperschaftsdatei)<sup>12</sup> sowie der inhaltlichen Erschließung nach RSWK (Regeln für den Schlagwortkatalog) und/oder DDC. Aufgrund der Masse der Netzpublikationen ist dies aber auf Dauer nicht mehr zu leisten. Die Deutsche Nationalbibliothek entwickelt derzeit aber ein neues, stärker automatisiertes Erschließungskonzept.

---

9 Informationen zum Metadaten-Kernset: [http://www.d-nb.de/netzpub/abliefe/pdf/metadaten\\_kernset\\_definitionen.pdf](http://www.d-nb.de/netzpub/abliefe/pdf/metadaten_kernset_definitionen.pdf)

10 Informationen zur Deutschen Nationalbibliografie: <http://www.d-nb.de/service/zd/dnb.htm>

11 Informationen zur Personennamendatei: <http://www.d-nb.de/standardisierung/normdateien/pnd.htm>

12 Informationen zur Gemeinsamen Körperschaftsdatei: <http://www.d-nb.de/standardisierung/normdateien/gkd.htm>

## Weitere Automatisierung der Geschäftsgänge

In einem nächsten Schritt ist die Automatisierung der Geschäftsgänge auf Seiten der Abnehmer geplant. Für die Ablieferung kleiner Mengen von Netzpublikationen ist das Webformular eine komfortable Lösung. Nach dem Einloggen in das Portal kann die Ablieferung erfolgen und sie dauert in der Regel auch nicht lange. Für die Massenablieferung wurde ein automatisiertes Harvestingverfahren<sup>13</sup> implementiert. Die Erstellung von Datensätzen erfolgt automatisiert ebenso wie die Verknüpfung mit der Netzpublikation. Es hat sich bereits im Umgang mit Online-Dissertationen gezeigt, welchen Vorteil einheitliche Metadatenstandards und Dateiformate bieten, insbesondere dann, wenn die Nachbearbeitung auf der Basis intellektueller Erschließungsinstrumente erfolgt. Deshalb sind Metadatenstandards erforderlich. Das Metadaten-Kernset bietet eine Konkordanz für den Import im ONIX-Format an. Weitere Einlieferformatstandards werden folgen, beispielsweise MARC21<sup>14</sup> und XMetaDiss(Plus).<sup>15</sup> Daneben sind individuelle Absprachen mit den Abnehmern erforderlich ebenso wie Testphasen, auch diese mit dem Ziel, die Automatisierung zu verbessern, mögliche Fehlerquellen bereits im Vorfeld auszuschalten und das Ausmaß notwendiger Datenprüfungen möglichst gering zu halten. Das Metadaten-Kernset für die Ablieferung von Zeitschriftenlieferungen ist in Vorbereitung. Hier wird beispielsweise in den Metadaten die Festlegung auf einen Identifier verlangt, über den die Heft- oder Artikellieferung mit dem Zeitschriftentitel verknüpft werden können.

Netzpublikationen stellen auch ein Mengenproblem dar. Die beschriebenen Verfahren sind ein erster Schritt hin zur notwendigen Automatisierung, zusätzliche Erweiterungen in diesem Bereich sind erforderlich, insbesondere in Bezug auf andere Dateiformate, aber auch auf weitere Ablieferungsverfahren wie z.B. das Webharvesting.

---

13 Informationen zum Verfahren der automatisierten Ablieferung über Harvestingverfahren finden sich auf der Website der Deutschen Nationalbibliothek unter [http://www.d-nb.de/netzpub/abliefer/pdf/automatisierte\\_ablieferung.pdf](http://www.d-nb.de/netzpub/abliefer/pdf/automatisierte_ablieferung.pdf) < urn:nbn:de:101-2009120311>

14 Informationen zum Umstieg auf MARC21: <http://www.d-nb.de/standardisierung/formate/marc21.htm>

15 Informationen zum Metadatenstandard XMetaDiss: <http://www.d-nb.de/standards/xmetadiss/xmetadiss.htm> . Allerdings ist hier eine Erweiterung auf weitere Hochschulschriften geplant. Bisher wurden nur die Online-Dissertationen gesammelt.

## Prinzipien der Langzeitarchivierung an der Deutschen Nationalbibliothek

Für die schnell wachsende Speichermenge muss nicht nur eine geeignete Datenverarbeitungs-Infrastruktur bereitstehen und gepflegt, gewartet und weiterentwickelt werden. Um die Inhalte der gesammelten Netzpublikationen über wechselnde Hard- und Softwaregenerationen zu bewahren, müssen die Daten mit geeigneten Metadaten in einem Archivsystem verwaltet werden, das die gängigen Langzeitarchivierungsstrategien unterstützt: Migration, die Konvertierung in aktuell nutzbare Dateiformate, und Emulation, die Herstellung von früheren Systemumgebungen auf aktuellen Systemen mit Hilfe spezifischer Software.

Die Langzeitarchivierung von Netzpublikationen an der Deutschen Nationalbibliothek basiert auf folgenden Prinzipien:

Grundsätzlich nimmt die Deutsche Nationalbibliothek Netzpublikationen in jedem Format an, die Ablieferer werden aber auf die Präferenzregelung hingewiesen (derzeit: 1. PDF/A, 2. Andere PDF-Versionen, 3. HTML, 4. PS, 5. Weitere XML-basierte Formate, TXT, 6. Sonstige (DVI, RTF, etc.).<sup>16</sup> Die Deutsche Nationalbibliothek archiviert nur eine von ggf. mehreren vorliegenden inhaltsgleichen elektronischen Dokumentversionen, wobei die Auswahl der Präferenzregelung folgt.

Jedes Objekt wird vor der Langzeitarchivierung automatisch mit technischen Metadaten angereichert, die den gezielten Zugriff auf Archivobjekte und die Anwendung von Langzeitarchivierungsstrategien unterstützen. Zur Analyse von Dateiformaten und zur automatischen Generierung von technischen Metadaten setzt die Deutsche Nationalbibliothek das Open Source Tool Jhove ein. Jhove (JSTOR/Harvard Object Validation Environment) ist ein Gemeinschaftsprodukt von JSTOR und der Harvard University Library (HUL). Jhove wird von einer großen internationalen Gemeinschaft benutzt, gepflegt und weiterentwickelt. Die Deutsche Nationalbibliothek bringt sich hier aktiv ein und arbeitet mit internationalen Partnern wie Harvard und der Niederländischen Nationalbibliothek (Koninklijke Bibliotheek) zusammen, z.B. an der Entwicklung fehlender Module für zusätzliche Formate.

Aus den abgelieferten und mit Metadaten versehenen Objekten werden unter Nachnutzung vorhandener Standards (z. B. METS) Archivobjekte im offenen definierten Paketformat Universelles Objektformat (UOF)<sup>17</sup> generiert. Dabei

---

16 Informationen zur Präferenzregelung: [http://www.d-nb.de/netzpub/ablief/np\\_dateiformate.htm](http://www.d-nb.de/netzpub/ablief/np_dateiformate.htm)

17 Informationen zum Universellen Objektformat:

kann ein Archivobjekt mehrere Dateien umfassen, die gemeinsam ein logisches Objekt, d.h. eine Netzpublikation ausmachen.

Die so entstandenen Archivobjekte werden in eine sichere Umgebung, das Archivsystem, eingespielt, in dem das gespeicherte Material ständig routinemäßig überprüft wird. Jedes Objekt wird durch Bitstream Preservation mit regelmäßigen Maßnahmen wie Backups und Umkopieren zur Sicherstellung der Datenintegrität unverändert im Originalformat erhalten. Die wichtigste Langzeitarchivierungsstrategie der Deutschen Nationalbibliothek ist zurzeit die Migration, denn große Mengen der Objekte, die unter den Sammelauftrag fallen, können damit adressiert werden. Wenn sich abzeichnet, dass die Originalformate zu veralten drohen, werden die archivierten Objekte in aktuelle, zukunftsfähige Formate migriert. Die Zielformate werden auf Grundlage kontinuierlicher Marktbeobachtung (Technology Watch) bestimmt. Bei Migrationen wird das Ausgangsobjekt immer erhalten und zusammen mit dem migrierten Objekt weiter aufbewahrt. Alle Migrationsschritte werden dokumentiert und in den Metadaten des Objekts verzeichnet.

### **kopal-Archivsystem**

Die technische Basis der Langzeitarchivierung an der Deutschen Nationalbibliothek bildet das Archivsystem, das im kopal-Projekt entwickelt wurde. Gefördert vom Bundesministerium für Bildung und Forschung (BMBF) hat die Deutsche Nationalbibliothek zwischen 2004 und 2007 in Partnerschaft mit der Niedersächsischen Staats- und Universitätsbibliothek Göttingen, der Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG) und der IBM Deutschland GmbH ein kooperativ nutzbares Langzeitarchiv aufgebaut.<sup>18</sup>

Das kopal-Archivsystem orientiert sich am OAIS-Referenzmodell (ISO 17421 „Open Archive Information System“) und setzt auf Standardsoftware auf. Für die Benutzung des kopal-Archivsystems entwickelten die Deutsche Nationalbibliothek und die SUB Göttingen die “kopal Library for Retrieval and Ingest” (koLibRI), die das Einspielen von Objekten in den Archivspeicher sowie den Zugriff auf die archivierten Objekte unterstützt.

Nach dem Abschluss der kopal-Projektphase 2007 erfolgt die Einbettung des Archivsystems in den Produktivbetrieb der Deutschen Nationalbibliothek. Die Einbettung erfordert einige konzeptionelle Anstrengung und Anpassungen in den vorhandenen technischen Systemen. Das Archivsystem wird so in die Geschäftsgänge eingebaut, dass die Netzpublikationen nahtlos von dem Zwi-

---

[http://kopal.langzeitarchivierung.de/index\\_objektspezifikation.php.de](http://kopal.langzeitarchivierung.de/index_objektspezifikation.php.de)

18 Informationen zu kopal: <http://www.kopal.langzeitarchivierung.de/>

schenspeicher, auf dem sie während des Erschließungsprozesses abgelegt sind, an das Archivsystem weitergegeben werden, wo sie langfristig und sicher aufbewahrt werden können. Um den Zugriff auf die Archivobjekte über Benutzerschnittstellen zu realisieren, müssen Schnittstellen angepasst werden. Weitere Schnittstellen müssen implementiert und Geschäftsgänge so umgestaltet werden, dass sie den Anforderungen der Archivobjekte gerecht werden. Zum Beispiel muss das Bereitstellungssystem darauf ausgerichtet werden, die archivierten Objekte im jeweils aktuellen Format (oder, alternativ, in dem vom Nutzer gewünschten Format) anzuzeigen. Auch die Anwendung von Langzeitarchivierungsstrategien im Praxisbetrieb, für die das Archivsystem ausgelegt ist, muss vorbereitet werden.

### **Weitere Herausforderungen**

Doch selbst das Zusammenspiel von bewährten Tools und sicheren Archivsystemen kann nicht alle Herausforderungen der Langzeitarchivierung lösen. Neben technischen müssen vor allen Dingen organisatorische Vorkehrungen getroffen werden, hier illustriert am Beispiel von Konvertierungseinstellungen von Dateiformaten. Im Prinzip kann das alle möglichen Formate betreffen, hier wird dies aber am Beispiel PDF erläutert, weil große Mengen der Archivbestände der Deutschen Nationalbibliothek in PDF vorliegen. Das Format ist bei Verlagen und anderen Ablieferern akzeptiert und weit verbreitet. Doch viele Verlage und Ablieferer liefern passwortgeschützte oder verschlüsselte PDFs ab oder deaktivieren bestimmte Funktionen wie zum Beispiel Druck- und Kopiermöglichkeiten. Das bereitet einerseits in der Benutzung der Dateien Probleme, wirft aber auch essentielle Probleme für die Langzeitarchivierung auf: An solchen Dateien können nicht alle Langzeitarchivierungsmaßnahmen durchgeführt werden und es können Datenverluste entstehen. Die Deutsche Nationalbibliothek ist daher im Gespräch mit Verlegern, um auf diese Problematik aufmerksam zu machen und für einheitliche, offene Speichereinstellungen zu werben. Gleichzeitig gilt es aber auch, die internen technischen Prozesse auf dieses Problem hin anzupassen: Entsprechende Dateien müssen zunächst automatisch erkannt und – unter Beachtung urheberrechtlicher Rahmenbedingungen – in eine für die Langzeitarchivierung geeignete Struktur überführt werden.

Um weitere Entwicklungen auf dem Gebiet der Langzeitarchivierung voranzutreiben, arbeitet die Deutsche Nationalbibliothek intensiv mit zahlreichen nationalen und internationalen Partnern zusammen. Dabei geht es sowohl um die zukünftige Anwendung von nötigen Langzeitverfügbarkeitsstrategien wie

Emulation und die kooperative Nutzung verschiedener Systeme (zum Beispiel in den EU-Projekten SHAMAN<sup>19</sup> und KEEP<sup>20</sup>), als auch um die Weiterentwicklung von Formatregistries wie GDFR und Pronom oder die gezielte Unterstützung der Entwicklung und breiten Anwendung von archivierungsfreundlichen Standards wie PDF/A.

## Ausblick

Die bestehenden Herausforderungen können insbesondere in Bezug auf die weiteren technischen Entwicklungen im Bereich der Netzpublikationen nur bewältigt werden, wenn die Verfahren verstärkt automatisiert werden. Dafür wurden die Grundlagen in der Systemarchitektur der Deutschen Nationalbibliothek gelegt. So können zumindest in Teilen die bereits entwickelten Verfahren für weitere Objekttypen nachgenutzt werden, aber es wird auch notwendig sein, neue Verfahren für Multimediaobjekte oder ablieferpflichtige Applikationen zu entwickeln.

Zum aktuellen Stand: Es werden Webformulare zur Ablieferung von Monografien, Zeitschriftenlieferungen (Hefte/Artikel) und Hochschulprüfungsarbeiten angeboten. Für die automatisierte Ablieferung über Harvestingverfahren sind Erweiterungen des Metadaten-Kernsets erforderlich, die u. a. auch für andere Objekttypen notwendig sein werden, wie beispielsweise für Audioobjekte. Weitere Entwicklungen betreffen die Metadatenformate, die in automatisierten Verfahren zur Anwendung kommen können, zum Beispiel auch die Anbindung weiterer Datenformate.

Eine Herausforderung stellt auch die Bereitstellung/Präsentation der Objekte dar. Erschwerend wirkt hier die rasche technologische Weiterentwicklung von Formaten und Abspielumgebungen. Übergeordnetes Ziel ist aber, die Bereitstellung für alle archivierten Objekte zu gewährleisten – auch für die Objekte, die auf einem Datenträger vorliegen. Angesichts der rund 700.000 Einheiten in den Sammlungen der Deutschen Nationalbibliothek ab ca. 1980 wird auch die historische Dimension dieses Problems offensichtlich.

---

19 Informationen zu SHAMAN: <http://www.d-nb.de/wir/projekte/shaman.htm> und <http://shaman-ip.eu/shaman/>

20 Informationen zu KEEP: <http://www.d-nb.de/wir/projekte/keep.htm> und <http://www.keep-project.eu/>



## 18.3 Langzeitarchivierung eines digitalen Bildarchivs – Projekt zum Aufbau eines Langzeitarchivs für hochaufgelöste digitale Bilddateien der Staatsgalerie Stuttgart am BSZ

Werner Schweibenz und Stefan Wolf<sup>21</sup>

*Der Beitrag beschreibt das Projekt der Staatsgalerie Stuttgart mit dem Bibliotheksservice-Zentrum Baden-Württemberg (BSZ) zum Aufbau eines Langzeitarchivs für hochaufgelöste digitale Bilddateien. Für die Archivierung wird das Langzeitarchiv SWBdepot des BSZ verwendet, die Metadaten für die Langzeitarchivierung werden mit dem Objektdokumentationssystem IMDAS-Pro erzeugt.*

### Die Situation in der Staatsgalerie Stuttgart

Die international bedeutsamen Sammlungen der Staatsgalerie Stuttgart (SGS)<sup>22</sup> reichen vom Tafelaltar bis zur modernen Medienkunst mit praktisch allen Formen z.B. der Malerei, Plastik, Graphik bis hin zu raumgreifenden Installationen. Sie werden fortlaufend erforscht, erschlossen und dokumentiert, gleichzeitig aber auch in verschiedenen Zusammenhängen eingebunden: beispielsweise in den eigenen Ausstellungen, im Leihverkehr zwischen Museen, in der Museumspädagogik, in Publikationen oder auch in Internetauftritten. Das Fotoatelier der SGS fertigt für diese Zwecke laufend eine große Zahl qualitativ hochwertiger Fotos an, dem Fortschritt der Technik folgend heute mit einer hochauflösenden Digitalkamera. Die anfallende Datenmenge wächst kontinuierlich. Für die Zukunft sucht die SGS nach einer kostengünstigen Lösung für Datenspeicherung und –sicherung bei gleichzeitiger Nutzung im Verbund. Dies gilt auch für die aus Sicherheitsgründen notwendige redundante Speicherung an einem zweiten Ort. Als Mitglied im MusIS-Verbund, dem landeseinheitlichen Verfahren für Museumsdokumentation der Staatlichen Museen in Baden-Württemberg wandte sich die SGS an das Bibliotheksservice-Zentrum Baden-Württemberg

---

21 Die Autoren danken herzlich den Kolleginnen und Kollegen der Staatsgalerie Stuttgart – allen voran Frau Dr. Elke Allgaier – für die kollegiale Unterstützung bei der Erstellung dieses Kapitels zum nestor Handbuch.

22 <http://www.staatsgalerie.de/>



BSZ),<sup>23</sup> das bereits für andere Institutionen ähnliche Dienstleistungen zu Dokumentmanagement und Langzeitarchivierung (LZA) anbietet.

### **Das Angebot des BSZ im Bereich Langzeitarchivierung**

Das BSZ bietet auf dem Gebiet der Langzeitarchivierung eine umfangreiche Dienstleistungspalette an.<sup>24</sup> Der Fokus des BSZ richtet sich auf die Verbindung der Objekte mit einer qualitativ hochwertigen Dokumentation und der Schaffung eines Mehrwerts für seine Kunden. Der Gewinn liegt z.B. in der technischen Realisierung spezifischer Geschäftsgänge für verschiedene Häuser auf gemeinsamer Basis, der gemeinsamen Nutzung der gleichen Software oder in der automatisierten Erzeugung und Bereitstellung von Gebrauchsderivaten aus den archivierten Objekten – also auf klassischen Synergieeffekten. Mit diesem Ansatz vereint das BSZ scheinbar Gegensätzliches, indem Produktion, Dokumentation und Nutzung mit der Archivierung wertvoller Daten in einen durchgängigen Arbeits- und Archivierungsprozess gebündelt werden. Für die Kunden wird ein zuverlässiges und kostensparendes Outsourcing großer Datenmengen in Kombination mit einer sicheren Datenhaltung der Originaldateien und einem flexiblen Zugriff auf Derivate geboten, also Mehrwerte, die eine Institution braucht, um nachvollziehbar, rationell und ökonomisch arbeiten zu können. Für die Bildproduktion der Museen erschien die Durchführung eines Pilotprojekts als notwendig, das vom Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg gefördert wird.

### **Die Leitidee**

Die Grundidee des Projektes zum „Aufbau eines Langzeitarchivs im BSZ für hochaufgelöste digitale Bilddateien der Staatsgalerie Stuttgart sowie die Entwicklung eines sicheren und auf Kontinuität basierenden Online-Daten-Transfers der Digitalisate“ (so der Name im Projektantrag) ist, unter strikter Beachtung des Primats einer sicheren und nachhaltigen Archivierung ein praktisches Verfahren zu entwickeln, bei dem ebenfalls nur ein Computersystem für die Dokumentation der zu archivierenden Bilddateien benötigt wird und bei dem

---

23 <http://www.bsz-bw.de/>

24 Wolf, Stefan ; Mainberger, Christof ; Schweibenz, Werner: Langzeitarchivierung am Bibliotheksservice-Zentrum Baden-Württemberg : Konzept, Aktivitäten und Perspektiven. – Preprint – Konstanz, BSZ, 2009.

URL: <http://opus.bsz-bw.de/swop/volltexte/2009/465/>

Erschienen in: Bibliotheksdienst, Heft 43(2009), Heft 3, S. 294-304.

die Mitarbeiter in allen beteiligten Arbeitsbereichen im Museum möglichst nur mit den ihnen bereits vertrauten Programmen und Werkzeugen arbeiten. Dieser Weg erhöht die Akzeptanz im Museum und hält den Einarbeitungs- und Schulungsaufwand gering. Gleichzeitig ist dies die Voraussetzung des zweiten Projektziels, das Verfahren nach Abschluss des Pilotprojekts den anderen Museen im MusIS-Verbund zur Verfügung zu stellen. Die Abläufe und Daten integrieren sich nahtlos in das Langzeitarchiv SWBdepot des BSZ. SWBdepot bezeichnet die am BSZ in Betrieb befindliche Speicherinfrastruktur, die nach Bedarf ausgebaut wird und auf der die üblichen Prozesse der Datensicherung wie z.B. Bandsicherung, redundante Speicherung und Konsistenzprüfungen abgewickelt werden.

## Der Produktionsablauf

Den Auftakt in der Produktion gibt die Bildbestellung, die in der Dokumentation festgehalten wird: nach ihrer Anweisung erstellt das Fotoatelier der SGS die Bildaufnahmen. Seit Januar 2005 arbeitet das Fotoatelier der SGS eingebettet in eine hausintern festgelegte Digitalisierungsstrategie mit einer Digitalkamera: Arbeitsprozesse, Geschäftsgänge und Dokumentation werden durchgängig elektronisch unterstützt und ausgeführt. Die professionellen Qualitätsansprüche an die Bildproduktion löst die digitale Fotografie mittlerweile ein. Je Museumsobjekt entsteht mindestens eine hochauflösende, unkomprimierte Masteraufnahme und ein farbkorrigierter, verlustfreier Submaster gleicher Auflösung im Tagged Image File Format (TIFF)<sup>25</sup> von durchschnittlich 50 MB bei 8 Bit Tiefe pro Farbkanal.

Das TIFF-Format bietet die Möglichkeit, IPTC-Metadaten<sup>26</sup> im Bild zu erfassen und zu speichern. Dieser Quasistandard der Pressefotografie erlaubt Angaben z.B. zu Bildrechten, Fotograf, Titel, Auflösung und Pixelzahl zu machen. Produkte wie z.B. Adobe Photoshop bieten dafür Erfassungsmasken, die im Fotoatelier von den Fotografen ausgefüllt werden. Teilweise handelt es sich um Angaben, die standardmäßig für jede Aufnahme aus der Staatsgalerie Stuttgart gemacht werden und in den Masken schon vorbelegt sind, teilweise aber auch um individuelle Merkmale, die zur einzelnen Fotografie eingetragen werden.

Sowohl Master als auch Submaster erhalten eindeutige Dateinamen, welche Hinweise auf Bildherkunft, Künstlernamen, Inventarnummer, Aufnahmegegebenheiten und Dateiformat enthalten. Zur Dateinamensgestaltung existiert eine formale und semantische Absprache zwischen SGS und BSZ. Die Konvention

---

25 Vgl.: [http://de.wikipedia.org/wiki/Tagged\\_Image\\_File\\_Format](http://de.wikipedia.org/wiki/Tagged_Image_File_Format)

26 Vgl.: <http://de.wikipedia.org/wiki/IPTC-NAA-Standard>

wird strikt eingehalten, da der Dateiname die Ablage im Archiv steuert und mit den IPTC-Daten zur Dokumentation herangezogen wird. Ideal ist es deshalb, wenn vor dem Transfer der Bilddatei das Objekt selbst schon als IMDAS-Pro-Museumsobjekt im Dokumentationssystem der SGS erfasst ist.

Vor dem sFTP-Transfer<sup>27</sup> der Dateien an das BSZ wird zur Sicherung der Integrität die MD5-Prüfsumme<sup>28</sup> berechnet. Die Tagesproduktion wird gebündelt und an das BSZ durch einen Cron-Job nachts transferiert. Alle folgenden Arbeitsprozesse werden zur Kontrolle der Datenintegrität mittels Vergleich bzw. Erhebung der MD5-Prüfsumme begleitet.

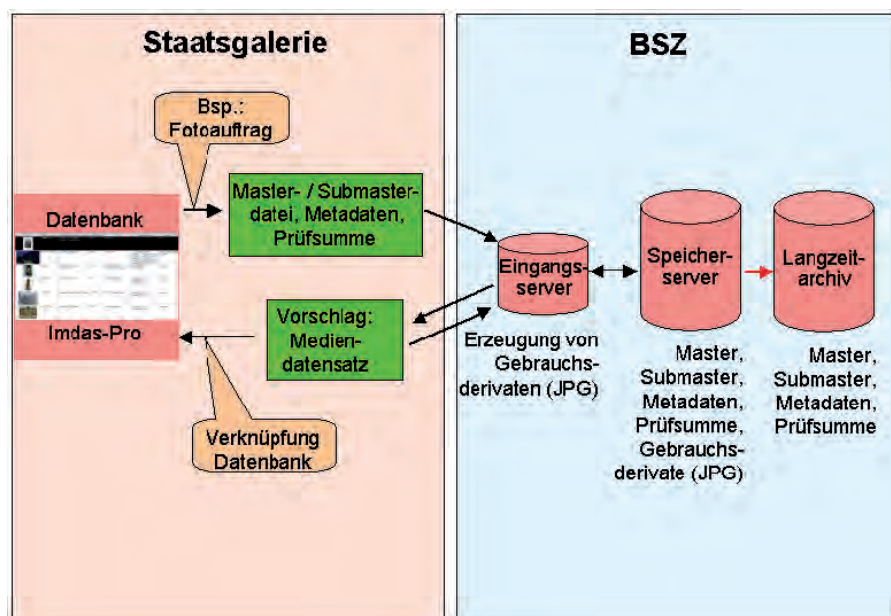


Abbildung 1<sup>29</sup>

27 [http://de.wikipedia.org/wiki/SSH\\_File\\_Transfer\\_Protocol](http://de.wikipedia.org/wiki/SSH_File_Transfer_Protocol)

28 [http://de.wikipedia.org/wiki/Message-Digest\\_Algorithm\\_5](http://de.wikipedia.org/wiki/Message-Digest_Algorithm_5)

29 Abbildung 1: Allgaier, Elke (Staatsgalerie Stuttgart): Archivierung von digitalen Bilddaten. Beitrag des nestor-Seminars „Digitale Langzeitarchivierung in Museen und Archiven - Konzepte und Strategien“ Köln, 21.11.2008. In: URL: [http://www.langzeitarchivierung.de/downloads/2008-11-21\\_allgaier.pdf](http://www.langzeitarchivierung.de/downloads/2008-11-21_allgaier.pdf)

Aufgabe des BSZ ist die Übernahme, Archivierung und Bereitstellung der Bilddaten. Nach Prüfung der formalen Eingangsvoraussetzung (Einhaltung der Dateinamenskonvention, korrekte Benennung des Dateiformats, gleichzeitige Lieferung von Master, Submaster und Prüfsumme etc.) werden die Bilder vom offenen FTP-Bereich in einen sicheren Arbeitsbereich kopiert. Der Master befindet sich danach in einem geschützten Archivbereich. Er wird nicht weiter verwendet, sondern bildet die Grundlage, wenn später auf die Ursprungsdatei der Bildaufnahme zurückgegriffen werden muss. Der Submaster wird physisch dupliziert – eine Kopie wird mit dem Master im Langzeitarchiv abgelegt, die zweite Kopie bereitgestellt und zur Erzeugung der Gebrauchsderivate im JPG-Format herangezogen. Nötig sind bislang jeweils ein kleines Thumbnail für die Vorschau in IMDAS-Pro und für die Präsentation in BAM, dem gemeinsamen Portal zu Bibliotheken, Archiven und Museen<sup>30</sup>, sowie eine größere Version zur differenzierten Betrachtung am Bildschirm oder für Restaurierung, Kunstvermittlung und entsprechende Zwecke.

Im gleichen Prozess werden Dateiname und IPTC-Daten gelesen und zusammen verarbeitet. Als Ergebnis entsteht eine Importdatei, die die nötigen Metadaten für eine Vorerfassung der Fotografie als Medienobjekt<sup>31</sup> in IMDAS-Pro bereitstellt. Sie enthält einen aus dem Dateinamen abgeleiteten Vorschlag zum Künstlernamen und zur Inventarnummer. Wegen der technischen Gegebenheiten in den Zeichensätzen kann die endgültige Ansetzung nicht transportiert werden. Die Importdatei enthält aber auch Angaben zu den technischen Daten der Fotografie und zu den Adressen bzw. Speicherorten der für die Nutzung bereitgestellten Derivate und Submaster, die aus den IPTC-Daten entnommen werden. Die Importdatei wird täglich in IMDAS-Pro eingelesen.

An dieser Stelle wird wieder die Staatsgalerie Stuttgart aktiv: die Dokumentation der Fotografie als IMDAS-Pro-Medienobjekt wird fertig gestellt und mit der Dokumentation des originalen Kunstwerks als IMDAS-Pro-Museumsobjekt verknüpft. Die Vorschläge zu Künstlernamen und Inventarnummer aus der importierten Vorerfassung werden in die endgültige, korrekte Ansetzungsform gebracht. Die Bilder werden geprüft und freigegeben, so dass nun auch im Dokumentationswerkzeug eine Vorausschau auf das Bild verfügbar ist und weitere Derivate sowie der Submaster per Mausklick nach Berechtigung angefordert werden können.

---

30 <http://www.bam-portal.de/>

31 Das IMDAS-Pro-Medienobjekt ist eine programminterne Dokumentationsklasse, die der Aufnahme von Objekten dient, die mit der Dokumentationsklasse Museumsobjekt in IMDAS-Pro verknüpft werden kann.

Die Verarbeitungsprogramme des BSZ sind modular verkettet, konfigurierbar und parametrisierbar. Weitere Gebrauchsderivate können auf Anforderung erzeugt, weitere Inhalte aus dem IPTC-Header ausgelesen werden. Zur Bildverarbeitung im BSZ ist die OpenSource-Software ImageMagick<sup>32</sup> integriert. Gleichzeitig überwacht das Programm auch die Abläufe, so dass Verantwortlichen an BSZ und SGS nach Abschluss des täglichen Jobs die nötigen Rückmeldungen erhalten.

IMDAS-Pro erlaubt den Export von Daten im XML-Format. Sie beinhaltet je nach Konfiguration sowohl Daten aus dem IMDAS-Pro-Medienobjekt als auch dem IMDAS-Pro-Museumsobjekt. Für die Langzeitarchivierung notwendig sind Metadaten in austauschfähigen, nicht proprietären Formaten – dafür bietet sich museumdat<sup>33</sup> bzw. das geplante Nachfolgeformat LIDO an (Lightweight Information Describing Objects, derzeit in der Entwicklungsversion 0.7 kursierend). Als Arbeitsergebnis der Fachgruppe Dokumentation des Deutschen Museumsbundes stellt es das Standardformat dar, mit dessen Hilfe Beschreibungsdaten aus Museumsbeständen ausgetauscht und gegenseitig nutzbar gemacht werden. Die aus IMDAS-Pro exportierten Daten werden in standardisierter Darstellung mit den Mastern und Submastern zu METS-Paketen verbunden und als Submission Information Packages an das Langzeitarchiv übertragen. In diese Pakete werden auch die notwendigen technischen Metadaten der Langzeitarchivierung eingebunden. Die automatische Erhebung dieser Daten geschieht mit der Open-Source-Software JHOVE, die am BSZ seit langem eingeführt ist und z.B. auch in kopal Anwendung findet. Im Archiv stehen diese Pakete den Prozessen der Langzeitarchivierung im engeren Sinn zur Verfügung.

## Der Projektstand

Die erste Arbeitssitzung im Projekt lag im April 2008. Im Frühjahr 2009 waren große Teile des Projektes einsatzbereit: die Vorgaben für Dateinamen und IPTC-Daten sind definiert; auf ihrer Basis wurden die Arbeitsrichtlinien in Dokumentation und Fotoatelier der Staatsgalerie erarbeitet, die notwendigen Schulungen fanden statt. Die Verarbeitungsprogramme im BSZ sind fertig gestellt und getestet, die Import-Routinen für IMDAS-Pro stehen bereit. Die notwendigen Anpassungsarbeiten an IMDAS-Pro der Version 4.0 wurden umgesetzt. Die Staatsgalerie Stuttgart hat einen großen Teil der seit 2005 herge-

---

32 <http://www.imagemagick.org/script/index.php>

33 XML-Schema und Dokumentation finden sich unter der URL <http://museum.zib.de/museumdat/>

stellten Fotografien mit mobiler Festplatte an das BSZ übertragen, die dort gesichert werden. Mit ihnen wurde ein Massentest der Verarbeitungsprogramme durchgeführt, der zeigte, dass die Programme zuverlässig arbeiten. Gleichzeitig steht mit dem Festplattentransport ein zweiter Lieferweg neben dem sFTP-Transfer zur Verfügung. Zusätzliche Anforderungen – z.B. an Reports aus den Prozessen – führen dazu, dass mit dem Produktionsbeginn im Frühjahr 2010 gerechnet wird. In Vorbereitung ist die Herstellung und Ablage der Submission Information Packages. Dafür wurde die Abbildung von museumdat auf METS/Premis im Rahmen eines Projektes untersucht. Mit Produktionsbeginn wird neben der Beachtung der nestor-Kriterien „Vertrauenswürdige digitale Langzeitarchive“<sup>34</sup> ein internes Audit zur Einhaltung der entsprechenden BSI-Kataloge samt Schutzbedarfsfeststellung durchgeführt.

### Die Perspektiven und Zusammenhänge

Nach gemeinsamer Einschätzung der Beteiligten hat das Projekt ein beträchtliches Potential im Hinblick auf Materialien, Anwendung und Ausbreitung. Das im Entstehen befindliche Verfahren entlastet die Staatsgalerie Stuttgart von der laufenden Server- und Softwarewartung für den digitalen Bildbestand und teilt die Verantwortlichkeit für die Langzeitarchivierung der Daten zwischen der SGS als Eigentümer der Daten und dem BSZ als Dienstleister. Neben positiven internen Organisationseffekten werden finanzielle Einsparungen erwartet. Das damit einhergehende fast vollständige Outsourcing des Datenbestandes ist für den Eigentümer der Originale und Originaldaten, die Staatsgalerie Stuttgart, nicht selbstverständlich, aber angesichts der Ergebnisse gewünscht, die Abläufe und Bedingungen zu erleichtern.

Gleichzeitig sind neben den hochauflösenden Fotografien eine Vielzahl weiterer, teilweise bereits historischer Bildbestände aus der Geschichte der Staatsgalerie, aus ihren Werkstätten, zu Veranstaltungen und Ausstellungen vorhanden, die für die Forschung immer relevanter werden. Auch wenn es sich vor allem um Aufnahmen im JPG-Format handelt, steht ihre langfristige Sicherung und Dokumentation an. Schon zur Vermeidung der Doppelerfassung bzw. paralleler Dokumentationswerkzeuge sind sie in die Langzeitarchivierung einzubeziehen.

Daneben entstehen auch Künstlervideos und weitere elektronische Inhalte

---

34 nestor-Kriterien - Kriterienkatalog vertrauenswürdige digitale Langzeitarchive Version II / hrsg. von der nestor-Arbeitsgruppe Vertrauenswürdige Archive - Zertifizierung. - Frankfurt am Main : nestor c/o Deutsche Nationalbibliothek, 2008. - 40 S.  
URN: urn:nbn:de:0008-2008021802

aus der Museumsdokumentation, aus Ausstellungsunterlagen, Öffentlichkeitsarbeit in Form von Texten, Bildern und Tonaufnahmen, die besondere Anforderungen an die Speichertechnik stellen. Ihre Bearbeitung steht in künftigen Projekten an.

Die in Deutschland maßgeblichen Standards der Langzeitarchivierung werden konsequent beachtet, proprietäre Lösungen vermieden. Die Standardisierung im Museumsbereich ist längst nicht so weit fortgeschritten wie z.B. im Bibliotheksbereich. Mit *museumdat* bzw. *LI-DO* steht ein Formatentwurf zur Verfügung, der den Verbindlichkeitsanforderungen an inhaltlich beschreibende Metadaten genügt und sich hoffentlich durchsetzen wird. Seine Benutzung bei der Bildung der zur Langzeitarchivierung üblichen *METS*-Objekte ist die zukunftsweisende Lösung. Solche Objekte werden geeignet sein, z.B. in eine Kopial-Installation übertragen zu werden.

Speicherplatz wird am BSZ bedarfsgerecht bereit gestellt und kann laufend erweitert werden. Damit verbunden sind die üblichen Sicherungsverfahren in einem Rechenzentrum inklusive dislozierter, redundanter Speicherung an einem zweiten Aufbewahrungsort.

Das Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg unterstützt das Projekt besonders auch aus einem weiteren Grund: nach Einführung des Verfahrens sollen die Projektergebnisse den anderen Museen des Landes Baden-Württemberg und des MusIS-Verbundes zur Verfügung stehen. Die Begleitung des Projekts durch das Badische Landesmuseum Karlsruhe (BLM) sichert genau diesen Sachverhalt: hausspezifische Lösungen werden vermieden, dafür werden Methoden und Wege gewählt, die sich mit geringem Aufwand verallgemeinern, übertragen und in weiteren Institutionen einführen lassen. Vereinbart ist schon heute, dass nach Inbetriebnahme der Produktion für die Staatsgalerie Stuttgart das BLM seinen Bestand an Fotografien im erarbeiteten Verfahren an das BSZ übertragen wird.

Es bestätigt sich, was in einem Workshop zur Langzeitarchivierung in Baden-Württemberg festgehalten wurde, an dem auf Einladung des Ministeriums für Wissenschaft, Forschung und Kunst des Landes Baden-Württemberg Vertreter von Rechenzentren, Bibliotheken, Archiven und Museen beteiligt waren und das seinen Niederschlag fand im Entwurf eines Schichtenmodells der Langzeitarchivierung digitaler Objekte<sup>35</sup>: der Bedarf an Lösungen für die Lang-

---

35 Vgl.: Wolf, Stefan: Das Schichtenmodell der digitalen Langzeitarchivierung in Baden-Württemberg : ein Konzeptpapier. Vortragsfolien vom 10. BSZ-Kolloquium am 21./22. September 2009 an der Hochschule der Medien, Stuttgart / Stefan Wolf. – Konstanz : BSZ, 2009. In: URL: <http://opus.bsz-bw.de/swop/volltexte/2009/771/>



zeitarchivierung in Museen besteht. Auch wenn die Digitalisierung der Bestände in den Museen später als z. B. in den Bibliotheken einsetzte und anderen Notwendigkeiten folgt, hat doch die Produktion von Daten, die verlässlich archiviert werden müssen, begonnen und wird einen beträchtlichen Aufschwung nehmen. Die Zusammenarbeit mit einem Dienstleister gewährleistet die gewünschte hohe Datensicherheit.

## Literatur

- Allgaier, Elke (Staatsgalerie Stuttgart): *Archivierung von digitalen Bilddaten*. Beitrag des nestor-Seminars „Digitale Langzeitarchivierung in Museen und Archiven - Konzepte und Strategien“ Köln, 21.11.2008. In: URL: [http://files.d-nb.de/nestor/veranstaltungen/2008-11-21\\_allgaier.pdf](http://files.d-nb.de/nestor/veranstaltungen/2008-11-21_allgaier.pdf)
- Wolf, Stefan / Mainberger, Christof / Schweibenz, Werner: *Langzeitarchivierung am Bibliotheksservice-Zentrum Baden-Württemberg : Konzept, Aktivitäten und Perspektiven*. – Preprint – Konstanz : BSZ, 2009. URL: <http://opus.bsz-bw.de/swop/volltexte/2009/465/> Erschienen in: *Bibliotheksdienst*, 43 (2009), Heft 3, S. 294-304.
- Wolf, Stefan: *Das Schichtenmodell der digitalen Langzeitarchivierung in Baden-Württemberg : ein Konzeptpapier*. Vortragsfolien vom 10. BSZ-Kolloquium am 21./22. September 2009 an der Hochschule der Medien, Stuttgart / Stefan Wolf. – Konstanz : BSZ, 2009. In: URL: <http://opus.bsz-bw.de/swop/volltexte/2009/771/>



## 18.4 ARNE – Archivierung von Netzressourcen des Deutschen Bundestages

*Angela Ullmann*

*Der Deutsche Bundestag archiviert seit Januar 2005 seine Internetangebote und stellt die archivierten Snapshots wiederum über ein Webarchiv im Internet bereit. Ausgehend von archivischen Grundprinzipien und den speziellen Rahmenbedingungen beim Deutschen Bundestag wurden für die Archivierung sowohl ein Konzept als auch ein System entwickelt. Dabei standen insbesondere Fragen zur Wahrung von Authentizität und Kontext, der archivischen Bewertung von Netzressourcen, aber auch technische Aspekte wie Maßnahmen zur Langzeiterhaltung im Fokus. Seit einiger Zeit kommen neue Herausforderungen hinzu – so die Wahrung des Persönlichkeitsschutzes beim freien Zugang zu archivierten Netzressourcen über das Internet.*

### **Fachliche Einordnung und organisatorische Rahmenbedingungen**

Die Archivierung von Netzressourcen des Deutschen Bundestages stellt ein Anwendungsbeispiel für ein fokussiertes Web-Harvesting dar. Die Sicherungsaufgabe leitet sich von der archivischen Zuständigkeit des Parlamentsarchivs ab und umfasst somit ausschließlich Webangebote, die aus der Provenienz „Deutscher Bundestag“ stammen. Konzeptionell einbezogen sind sowohl öffentlich zugängliche Angebote im Internet als auch nichtöffentliche wie das Intranet.<sup>36</sup> In den Wirkbetrieb überführt ist bislang nur die Archivierung einzelner Internetangebote, die zum Zeitpunkt ihrer Veröffentlichung bzw. des Downloads keinen Zugangsbeschränkungen unterlagen.

Die Archivierung wird in Kooperation zwischen zwei Organisationseinheiten der Bundestagsverwaltung realisiert: dem Referat „Parlamentsarchiv“ und dem Referat „Online-Dienste / Parlamentsfernsehen“. Die Arbeitsteilung beruht auf den verwaltungsmäßigen Zuständigkeiten und den daraus resultierenden Kompetenzen. Während das Parlamentsarchiv die archivische Bewertung und die anderen anfallenden (archiv)fachlichen Aufgaben wahrnimmt, wird die technische Abwicklung durch die Online-Dienste übernommen. Da

---

36 Das Intranetangebot des Deutschen Bundestages ist im Unterschied zum Internet nur für einen beschränkten Adressatenkreis zugänglich. Zu diesem gehören die Abgeordneten und deren Mitarbeiter sowie die Mitarbeiter der Fraktionen und der Bundestagsverwaltung.

auch die inhaltliche und technische Pflege der aktuellen Webangebote dort resortiert, sind Informationsverluste oder Kommunikationslücken zu anstehenden Veränderungen der Webangebote weitgehend ausgeschlossen.

## **Entwicklung und Fortschreibung einer Archivierungslösung**

Basierend auf Vorüberlegungen des Parlamentsarchivs zur Sicherung von Netzressourcen des Deutschen Bundestages aus den Jahren 2002 bis 2003 wurde im Jahre 2004 eine Übereinkunft mit den Online-Diensten zur Entwicklung einer Archivierungslösung für das Angebot „Bundestag im Internet“ ([www.bundestag.de](http://www.bundestag.de)) getroffen. Der enge Fokus war die einzige Möglichkeit, sich dieser Aufgabe praxisnah zu nähern. Gleichzeitig liegt hier jedoch ein „Geburtsfehler“ des „Systems zur Archivierung von Netzressourcen des Deutschen Bundestages“ (ARNE), da es systemtechnisch eng an eine einzelne Ressource gebunden ist. Da diese Ressource jedoch das wichtigste Angebot des Bundestages im Netz ist, das unbedingt erhalten werden soll, war dieses Vorgehen gerechtfertigt.

Standards blieben in der Entwicklungsphase weitgehend unberücksichtigt. Zum einen wurde und wird der (Meta)Datenaustausch mit anderen Gedächtnisorganisationen nicht angestrebt, daher musste keine Evaluierung eventuell geeigneter Metadatenstandards erfolgen. Auch für die Zugrundelegung des OAIS-Modells gab es beim Start des Vorhabens kein Bedürfnis.

In Abweichung zur üblichen Verfahrensweise bei derartigen Projekten entstanden Konzept und System zeitgleich und in gegenseitiger Abhängigkeit. Die wenigen bislang existierenden Referenzprojekte waren für die beim Bundestag angestrebte Lösung nicht einschlägig – entweder hinsichtlich der archivfachlichen Prämissen oder der Einbettung in bestehende Systemlandschaften. Das weitaus bekannteste Referenzprojekt dürfte das Internet Archive (<http://web.archive.org>) sein. Eine Analyse der dort gespeicherten Snapshots (also Momentaufnahmen) von „Bundestag im Internet“ ergab, dass durch die Wayback-Machine Internetseiten verschiedener Zeitschnitte miteinander in einem Angebot verbunden werden, die vor der Archivierung nicht gleichzeitig online verfügbar waren. Der Benutzer erhält jedoch beim Laden der Seiten keine entsprechenden Hinweise. Dies verstößt eindeutig gegen den Grundsatz der Authentizität. Die Wahrung der Authentizität ist eine große Herausforderung, aber auch eine entscheidende Frage nicht nur bei der Webarchivierung.

Die während der Archivierungs- und Aufbereitungsvorgänge aufgetretenen Fehler und Fragen ermöglich(t)en das (Fort)Schreiben eines Konzeptes, das der Wirklichkeit entspricht. Diese induktive Methode empfiehlt sich zum jetzigen

Zeitpunkt insbesondere für kleinere Einrichtungen und einzellige Archive, deren Sicherungsauftrag nur wenige Webangebote umfasst.

Generell ist davon auszugehen, dass Systeme zur Webarchivierung einem unablässigen Wandel unterworfen sind, da sich auch die Technologien zur Erzeugung von Webangeboten rasant weiterentwickeln.

## Grundsätze und Anforderungen

Die Bewahrung aller Webseiten in allen jemals veröffentlichten Versionen wurde bereits zu Beginn der internen Diskussion weder als realistisch noch als sinnvoll angesehen. Die Archivierung erfolgt daher auf zwei Wegen: einmal als Turnus- und darüber hinaus als Anlassarchivierung – auch bezeichnet als selective Harvesting und Eventharvesting<sup>37</sup>, wobei diese Methoden nicht gleichberechtigt nebeneinander stehen, sondern das Eventharvesting eine Nebenform des zyklischen selektiven Harvesting darstellt.

Der Zyklus der Archivierung unterscheidet sich für jede Netzressource. Die dabei berücksichtigten Aspekte sind unter „Auswahlstrategie und Bewertung der Netzressourcen“ ausgeführt.

Bewahrenswert sind aus archivischer Sicht nicht nur die Informationen, die auf einer Webseite publiziert sind, sondern auch das Aussehen, die Gestaltung, die Funktionalitäten, der Kontext und das Verhalten der Webseiten. Als Bezugspunkt dient dabei die Nutzersicht und nicht die Sicht des Systembetreuers, der auch auf Inhalte zugreifen kann, die aktuell nicht freigegeben sind. Somit finden bei der Archivierung nur die zu diesem Zeitpunkt veröffentlichten Seiten / Dateien Berücksichtigung. Diese Entscheidung gab den Ausschlag für die Wahl eines Crawlers und gegen FTP für den Downloadprozess.

Ein weiterer Grundsatz bestand im Bestreben, archivierte Netzressourcen möglichst zeitnah wiederum dem ursprünglichen Adressatenkreis auf einem gleichwertigen Zugangsweg wie vor der Übernahme in das Archiv bereitzustellen.

Damit unmittelbar verbunden war die Frage, ob alle im vorarchivischen Bereich angebotenen Funktionen (Druckversion erzeugen, mailto-Befehle etc.) im Webarchivsystem nachzubilden sind, ob dies vom Aufwand her vertretbar und auch sinnvoll ist. Aus unterschiedlichen Gründen wurde entschieden, diese Funktionen nicht anzubieten: Der mailto-Befehl bspw. soll nicht mehr ausführbar sein; das Erzeugen der Druckfunktion war technisch zu aufwändig. Letz-

---

37 Vgl. PoWR. *The Preservation of Web Resources Handbook. Digital preservation for the UK HE/FE web management community*. London 2008. URL <http://www.jisc.ac.uk/media/documents/programmes/preservation/powrhandbookv1.pdf>, S. 18 - 19

teres hat auch Kritik von Nutzern hervorgerufen, weil die Erwartungen an die Funktionalität archivierter Netzressourcen die gleichen sind wie die an Live-Angebote im Web.

Für die Erhaltung interaktiver Inhalte wie bspw. den virtuellen Adler auf „Bundestag im Internet“ oder das Einreichen einer elektronischen Petition ist bislang noch keine Lösung entwickelt. Aus archivischer Sicht wäre eine Bewahrung dieser Angebote unstreitig wünschenswert, während andere interaktive Inhalte wie bspw. die Bestellung von Informationsmaterial zum Deutschen Bundestag als nicht archivwürdig gelten können. Bislang bleiben alle interaktiven Inhalte von der Archivierung ausgeschlossen.

### **Auswahlstrategie und Bewertung der Netzressourcen**

Die Bewahrung von Webangeboten des Deutschen Bundestages erfolgt nach archivischen Prinzipien. Ein grundlegendes Prinzip ist die unter „Fachliche Einordnung und organisatorische Rahmenbedingungen“ bereits erläuterte Provenienzbindung. Ein weiterer Grundsatz besteht in der Auswahl der zu archivierenden Unterlagen. Dieser findet auch auf Webangebote Anwendung. Die Webprojekte des Bundestages werden in ihrer Entstehung und Entwicklung beobachtet und bewertet. Die Bewertungsentscheidung ist zweistufig: zunächst wird entschieden, welche Netzressourcen grundsätzlich archivwürdig sind. Die Archivwürdigkeit wurde mit einer Ausnahme für alle aktuellen Webprojekte des Bundestages bejaht. Die Entscheidung, die Homepage „Das Parlament“ ([www.das.parlament.de](http://www.das.parlament.de)) nicht zu archivieren, beruht auf dem fehlenden inhaltlichen Mehrwert, denn sie stellt eine nahezu identische Webaufbereitung der vom Deutschen Bundestag im Druck herausgegebenen Wochenzeitung „Das Parlament“ dar. Die Darbietung im Netz ist lediglich ein alternativer Verbreitungsweg. Aus archivischer Sicht handelt es sich somit um eine Mehrfachüberlieferung.

Im positiven Falle ist darüber hinaus das Intervall einer Archivierung festzulegen. Diese archivische Bewertungsentscheidung orientiert sich an der Aussagekraft und dem Stellenwert der Netzressource, dem Aktualisierungsintervall des Live-Angebotes, den inhaltlichen Alleinstellungsmerkmalen des Webangebotes und dem Ziel der Archivierung. Das Webangebot „Mitmischen“ ([www.mitmischen.de](http://www.mitmischen.de)) für Jugendliche behandelt allgemeine aktuelle politische Themen und soll Jugendliche an Politik heranführen. Der inhaltliche Bezug zum Deutschen Bundestag ist hier nicht unmittelbar gegeben. Bei diesem Angebot ist ein halbjährlicher Archivierungsturnus vorgesehen. Das Ziel der Archivierung besteht dabei nicht in der Nachvollziehbarkeit aller Informationen über politische Ereignisse, sondern einer auswahlweisen Veranschaulichung des Angebotes.

„Bundestag im Internet“ wird dagegen regelmäßig alle vier Wochen archiviert. Dieses Intervall kann in einer „heißen“ politischen Phase geändert werden. Die politischen Ereignisse im Jahre 2005 mit der Vertrauensfrage des Bundeskanzlers, der Verkürzung der Legislaturperiode und der vorzeitigen Neuwahl waren Anlass dafür, das Archivierungsintervall in dieser Zeit auf 14 Tage festzulegen.

Politische Ereignisse können ebenso wie technisch-inhaltliche Veränderungen (Redesign des Webangebotes o.ä.) eine zusätzliche Anlassarchivierung außerhalb des normalen Archivierungsturnus nach sich ziehen.

Die Grundsätze für die Auswahl und Bewertung sind durch die ständig aktualisierte Veröffentlichung des Konzeptes zur „Archivierung von Netzressourcen des Deutschen Bundestages“<sup>38</sup> für jedermann transparent.

### **Wahrung der Authentizität, Erschließung und Metadaten**

Das oberste Gebot der Archivierung ist die Wahrung der Authentizität. Ein Dokument muss immer das sein, was es zu sein vorgibt. Die konsequente Anwendung dieses Prinzips gestattet durchaus Veränderungen an einer Netzressource im Zuge ihrer Archivierung – allerdings müssen alle Änderungen dokumentiert und jederzeit durch jedermann nachvollziehbar sein. Realisieren lässt sich dies vorrangig durch Metadaten.

Metadaten dienen nicht nur zur Dokumentation der technischen Ursprungs-umgebung sowie der mit einer Archivierung verbundenen technischen Maßnahmen und deren Parameter. Sie geben auch inhaltliche Auskünfte und erschließen eine Netzressource – beispielsweise durch die Angabe der Provenienz, des Archivierungsdatums, des Archivierungsanlasses, der Domäne.

Für ARNE wurde eine Liste von Metadaten definiert, die sowohl den vorarchivischen Bereich als auch den Workflow der Archivierung beschreibt. Diese Liste wird im Rahmen der Fortentwicklung des Systems ebenfalls erweitert. Die Aufzählung aller Metadaten würde den Rahmen dieses Beitrages sprengen. Es soll daher an dieser Stelle abermals auf die Konzeption zur „Archivierung von Netzressourcen des Deutschen Bundestages“ verwiesen werden.

Weitere Gesichtspunkte für die Wahrung der Authentizität von Netzressourcen sind die Behandlung externer Links und die Art der Aufbereitung für eine Nutzung. Die Verlinkung von Webseiten gehört zur elementaren Charakteristik dieser Quellengattung. Links verbinden sowohl Inhalte innerhalb einer Netzressource, führen aber auch zu anderen Webangeboten entweder desselben Inhabers oder zu Angeboten Dritter. Bei der Archivierung von Netzressourcen des

---

38 zur URL siehe Ende des Beitrages

Deutschen Bundestages bleiben nur die Links innerhalb eines Webangebotes im archivierten Snapshot unmittelbar ausführbar. Absolute Links werden hierzu im Rahmen der Konvertierung in interne Links umgewandelt. Für alle anderen Links werden die Ziele und das Verhalten („öffne ein neues Fenster“ etc.) gesichert.

Auf die Sicherung der Authentizität im Rahmen der Aufbereitung für eine Nutzung wird unter „Bereitstellung und Nutzung“ eingegangen.

## **Workflow**

Erst über die Jahre des Wirkbetriebs hinweg konnte der Workflow zu einer stimmigen Abfolge entwickelt werden. Aktuell besteht er aus den Arbeitsschritten:

- Archivische Bewertung aller Netzressourcen
- Registrierung technischer Metadaten zu potentiellen Dateiformaten, die in einer Netzressource des Bundestages enthalten sein können
- Technische Verankerung der Archivierungsoptionen
- Anlegen eines Snapshots in der Referenzdatenbank (damit automatisch verbunden Anlegen eines Verzeichnisses für die Ablage des Snapshots auf dem Webarchivserver)
- Start und Ablauf des Downloadvorganges
- Anlegen einer Kopie für den gesamten Snapshot
- Konvertierung (umfasst mehrere Arbeitsschritte wie Umwandlung der externen Links, Konvertierung der html-Seiten nach xhtml)
- Indexierung
- Qualitätssicherung durch die Prüfung von definierten Referenzseiten
- Freigabe für die Benutzung (damit automatisch verbunden Transfer auf den externen Webserver und Bereitstellung im Internet)
- Backup
- Weitere Erhaltungsmaßnahmen

In diesen Workflow eingebunden ist die Ermittlung der technischen Zusammensetzung einer Netzressource oder auch neu hinzugekommener Dateiformate.

## **Bereitstellung und Nutzung**

Netzressourcen des Bundestages sollen so bald als möglich nach ihrer Archivierung im Internet verfügbar gemacht werden. Es existieren mehrere Zugangswege:

- Das Webarchiv ist eingebunden in „Bundestag im Internet“.
- Von verschiedenen Seiten in „Bundestag im Internet“ wird auf ältere Inhalte im Webarchiv direkt verlinkt.
- Das Webarchiv ist über Suchmaschinen zugänglich.
- Jeder Nutzer kann einen Link direkt auf eine archivierte Webseite erzeugen und diesen auf die übliche Art verwenden.
- Künftig soll es auch über die Archivdatenbank des Parlamentsarchivs erreichbar sein.

Archivierte Snapshots müssen jederzeit als Archivgut erkennbar sein. Der Nutzer muss sehen, dass er sich im Archiv befindet und um welchen Snapshot es sich handelt. Auch ein Wechsel in einen anderen Snapshot muss erkenn- und nachvollziehbar sein.

Alle archivierte Webseiten des Bundestages werden daher in einem roten Rahmen dargestellt. Die Kopf- und Fußzeile dieses Rahmens informieren über den Archivstatus und zeigen die wichtigsten Metadaten zur Identifizierung des Snapshots an.

Bei der Ausführung von Links, die aus dem Snapshot hinausführen („externe Links“) unterscheidet das System verschiedene Arten und gibt in Abhängigkeit davon Hinweise für den Benutzer. Bei externen Links zu Datenbanken oder Angeboten Dritter erscheint ein Hinweis auf den „verlorenen“ Kontext. Dem Benutzer wird erklärt, dass die referenzierte Datenbank bzw. das referenzierte Webangebot seit der Archivierung der genutzten Netzressource wahrscheinlich inhaltlich und gestalterisch verändert wurde. Damit kann nicht (mehr) von einem unmittelbaren Bezug der Netzressource auf die referenzierte Datenbank oder das referenzierte Webangebot ausgegangen werden („Verlorener Kontext“). Bei Links in andere Snapshots erhält der Benutzer eine Mitteilung darüber, dass er in einen anderen Snapshot wechselt.

Eine besondere Herausforderung stellen Inhalte dar, die nach ihrer Archivierung aufgrund datenschutzrechtlicher oder anderer Bestimmungen in der Benutzungsversion geändert werden müssen. In diesem Rahmen vorgenommene Änderungen sind zur Wahrung der Authentizität zu dokumentieren. Eine fachliche Lösung und deren technische Umsetzung hierfür werden im Rahmen von ARNE angestrebt und befinden sich derzeit in Entwicklung.

## **Maßnahmen zur langfristigen Erhaltung**

Verschiedene Maßnahmen sollen zur langfristigen Erhaltung beitragen, auch wenn ein umfassendes Sicherungskonzept noch aussteht. Die technischen und



inhaltlichen Metadaten geben ein umfassendes Bild der Netzressource. Bestandteil dieser Metadaten sind auch Fehlerprotokolle und Logbücher der archivtechnischen Bearbeitung sowie Dateistatistiken.

Nach dem Download werden die Daten zunächst kopiert, um jederzeit auf die unbearbeitete Version des Snapshots zurückgreifen zu können. Konvertiert werden bislang lediglich html-Dateien nach xhtml.

Darüber hinaus existiert ein Datensicherungskonzept, in das interne und externe Datenträger an unterschiedlichen Standorten einbezogen sind.

### **Technische Eckpunkte**

Beim System ARNE handelt es sich um eine Eigenentwicklung des Deutschen Bundestages. Es umfasst u.a. die Referenzdatenbank für die Metadaten und verbindet verschiedene am Markt erhältliche Tools bspw. für den Download (httrack) und die Indexierung.

Die Speicherung der Daten erfolgt auf einem gesonderten Webarchivserver, der sich bei den Online-Diensten befindet und dort auch technisch betreut wird.

Eine detaillierte technische Beschreibung ist dem Konzept zur Archivierung von Netzressourcen zu entnehmen.

### **Mehrwert der Webarchivierung für den Deutschen Bundestag**

Was aber bedeutet die Archivierung von Netzressourcen für den Deutschen Bundestag? War es zunächst ausschließlich ein archivisches Anliegen, historisch wertvolle Quellen zu bewahren, so erbringt die Webarchivierung mittlerweile wichtige institutionelle Mehrwerte.

Bereits die Beschäftigung mit den Webangeboten im Rahmen der Vorbereitung der Webarchivierung brachte wichtige Erkenntnisse und Fragen auf die Tagesordnung. Das System zur Webarchivierung offenbarte technische Fehler im Live-Angebot, die so vorher nicht ersichtlich waren. Allein die aus nicht mehr zielführenden Links resultierenden „Fehlerseiten“ konnten in den letzten Jahren erheblich reduziert werden.

Das Webarchiv hat sich zu einem wichtigen institutionellen Gedächtnis entwickelt. So lässt sich nachweisen – wenn auch nicht lückenlos –, welche Informationen zu welcher Zeit online verfügbar waren. Es trägt damit nicht zuletzt zur Rechtssicherung bei.

Durch die regelmäßige Archivierung von Webseiten wird das vorarchivische Content-Management-System (CMS) entlastet. Ältere Inhalte, die nicht mehr fortgeschrieben werden, können im CMS gelöscht und aus dem aktuellen



Angebot ins Webarchiv verlinkt werden. Ältere Seiten müssen dann auch nicht mehr in Relaunches einbezogen werden.

## Literatur und Quellenangaben

*Ausführliche Darstellung in:*

Ullmann, Angela / Rösler, Steven (2007): *Archivierung von Netzressourcen des Deutschen Bundestages*. Version 2.0. In: Online-Veröffentlichungen aus dem Parlamentsarchiv des Deutschen Bundestages. Dezember 2007. [http://www.bundestag.de/wissen/archiv/oeffent/arch\\_netz\\_gross2.pdf](http://www.bundestag.de/wissen/archiv/oeffent/arch_netz_gross2.pdf)

PoWR. *The Preservation of Web Resources Handbook. Digital preservation for the UK HE/FE web management community*. London 2008., S. 18 – 19. URL <http://www.jisc.ac.uk/media/documents/programmes/preservation/powrhandbookv1.pdf>

*Webarchiv des Deutschen Bundestages im Internet:*

<http://webarchiv.bundestag.de>

*Rechercheanleitung für das Webarchiv:*

[http://webarchiv.bundestag.de/cgi/recherche\\_anleitung\\_webarchiv\\_bundestag.pdf](http://webarchiv.bundestag.de/cgi/recherche_anleitung_webarchiv_bundestag.pdf)