H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, M. Jehn (Hrsg.)

nestor Handbuch

Eine kleine Enzyklopädie der digitalen Langzeitarchivierung

Version 2.0

Kapitel 18.2
Langzeitarchivierung von elektronischen Publikationen durch die Deutsche Nationalbibliothek

nestor

vwh

nestor Handbuch: Eine kleine Enzyklopädie der digitalen Langzeitarchivierung
hg. v. H. Neuroth, A. Oßwald, R. Scheffel, S. Strathmann, M. Jehn
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
c/o Niedersächsische Staats- und Universitätsbibliothek Göttingen,
Dr. Heike Neuroth, Forschung und Entwicklung, Papendiek 14, 37073 Göttingen

Die Herausgeber danken Anke Herr (Korrektur), Martina Kerzel (Bildbearbeitung) und Jörn Tietgen (Layout und Formatierung des Gesamttextes) für ihre unverzichtbare Unterstützung bei der Fertigstellung des Handbuchs.

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.

Die Inhalte dieses Buchs stehen auch als Onlineversion (http://nestor.sub.uni-goettingen.de/handbuch/) sowie über den Göttinger Universitätskatalog (http://www.sub.uni-goettingen.de) zur Verfügung.

www.vwh-verlag.de
In Kooperation mit dem Universitätsverlag Göttingen

Druck und Bindung: Kunsthaus Schwanheide

Printed in Germany – Als Typoskript gedruckt –

ISBN: 978-3-940317-48-3

URL für Kapitel 18.2 „Langzeitarchivierung von elektronischen Publikationen durch die Deutsche Nationalbibliothek“ (Version 2.0): urn:nbn:de:0008-20090811974
http://nbn-resolving.de/urn/resolver.pl?urn=urn:nbn:de:0008-20090811974

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

18.2 Langzeitarchivierung von elektronischen Publikationen durch die Deutsche Nationalbibliothek

Maren Brodersen und Sabine Schrimpf [1]

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".[2] 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://www.d-nb.de/wir/pdf/dnbg.pdf
Alle hier aufgeführten URLs wurden im April 2009 auf Erreichbarkeit geprüft .

In der Pflichtablieferungsverordnung vom 17. Oktober 2008 (PflAV)[3] 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.[4]

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 werden unter http://www.d-nb.de/netzpub/index.htm veröffentlicht werden.

verbunden. Für die Deutsche Nationalbibliothek verändert sich damit die Selektion der Ablieferer von Netzpublikationen. In der Printwelt sind die Ablieferer 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)[5] gemeldet. Teilweise sind die Ablieferer von Netzpublikationen identisch mit den bereits bekannten Ablieferern 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 Ablieferer 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 Ausgabenformate 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[6] 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.[7] 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 harmonieren.

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[8] 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 Identifier

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

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

(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 Artikelebene. Im Formular erfolgt die Verknüpfung über den in einer Auswahlliste angezeigten Zeitschriftentitel. Damit können Zeitschriften auf Heft- oder Artikelebene recherchiert und anzeigt 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[9] 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[10] 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)[11] und GKD (Gemeinsame Körperschaftsdatei)[12] 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/info/pdf/metadaten_kernset_extern.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 Ablieferer 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 wird derzeit gerade ein automatisiertes Harvestingverfahren[13] mit einem Verlag getestet. Auch hier ist die Voraussetzung, dass die Erstellung von Datensätzen automatisiert erfolgt 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 standartisierte Metadaten erforderlich. Das Metadaten-Kernset bietet eine Konkordanz für den Import im ONIX-Format an. Weitere Einlieferformatstandards werden folgen, beispielsweise MARC21[14] und XMetaDiss(Plus).[15] Daneben sind individuelle Absprachen mit den Ablieferern 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/ablief/pdf/automatisierte_ablieferung.pdf

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 (09.02.2009). Allerdings ist hier eine Erweiterung auf weitere Hochschulschriften geplant. Bislang 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.).[16] 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 offen definierten Paketformat Universellen Objektformat (UOF)[17] generiert. Dabei

16 Informationen zur Präferenzregelung: 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.[18]

Das kopal-Archivsystems 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

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[19] und KEEP[20]), 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: Die Formulare zur Ablieferung von Zeitschriftenlieferungen (Hefte/Artikel) sind in der Testphase. 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

20 Informationen zu KEEP: http://cordis.europa.eu/fetch?CALLER=PROJ_ICT&ACTION=D&DOC=13&CAT=PROJ&QUERY=011f22fab3a8:cf13:10a93ccb&RCN=89496