

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

nestor Handbuch

Eine kleine Enzyklopädie
der digitalen Langzeitarchivierung

Version 2.3

Kapitel 18.2

Langzeitarchivierung von
elektronischen Publikationen durch
die Deutsche Nationalbibliothek

nestor 

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
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 2.3.1 ist eine Printversion 2.0 beim Verlag Werner
Hülsbusch, Boizenburg erschienen.

Die digitale Version 2.3.1 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 18.2 „Langzeitarchivierung von elektronischen Publikationen durch die
Deutsche Nationalbibliothek“ (Version 2.3.1): [urn:nbn:de:0008-2018040508](http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2018040508)
<http://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:0008-2018040508>



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 Netzpublikationen durch die Deutsche Nationalbibliothek

Dr. Dirk Weisbrod¹

Abstract

Seit 2006 sammelt die Deutsche Nationalbibliothek auch im gesetzlichen Auftrag „Medienwerke in unkörperlicher Form“ und damit Netzpublikationen. Um der Erweiterung des Sammelauftrages gerecht zu werden, musste die DNB eine Langzeitarchivierungs-Strategie und entsprechende Geschäftsprozesse entwickeln sowie die dafür nötige technische Infrastruktur aufbauen. Auch die Pflichtablieferung der neuen Sammelobjekte musste modelliert werden. Der nachfolgende Beitrag schildert den aktuellen Stand der Langzeitarchivierung in der DNB und gibt auch einen Ausblick auf zukünftige Entwicklungen.

Sammelgebiet

Der Sammelauftrag der Deutschen Nationalbibliothek (DNB) wird vom Gesetz über die Deutsche Nationalbibliothek (DNBG) vom 22.06.2006 und der zuletzt am 29. April 2014 geänderten Verordnung über die Pflichtablieferung von Medienwerken an die Deutsche Nationalbibliothek (Pflichtablieferungsverordnung - PflAV) definiert und von den Sammelrichtlinien der DNB (derzeitiger Stand 01.05.2014) konkretisiert. Mit dem Inkrafttreten des DNBG wurden 2006 auch Medienwerke in unkörperlicher Form und damit Netzpublikationen² in den Sammlungsauftrag der DNB aufgenommen. Sie sind ebenso wie ablieferungspflichtige digitale Medien auf Datenträgern „zu bewahren und ihre Verfügbarkeit und ihre Nutzbarkeit langfristig, d.h. für eine unbestimmte Zeitspanne und über technologische Veränderungen hinaus, zu erhalten.“³ Dabei erfolgt der Zugriff in der Regel in den Lesesälen der DNB. Wenn eine entsprechende Lizenzierung vorliegt ist der Online-Zugriff über das DNB-Portal möglich.

Bei der Sammlung von Netzpublikationen unterscheidet die DNB zwischen „Verlagsveröffentlichungen, wissenschaftlichen, institutionellen und kulturellen

¹ Mit Unterstützung von Susanne Puls, Karlheinz Schmitt, Sabine Schrimpf und Tobias Steinke

² Die DNB definiert eine Netzpublikation als ein „unkörperliches Medienwerk, das der Öffentlichkeit zugänglich gemacht wurde. Synonym dazu wird häufig der Begriff Online-Publikation verwendet.“ Aus der Sicht der Langzeitarchivierung sind Netzpublikationen digitale Objekte, die dauerhaft zugänglich gemacht werden müssen.

³ Sabine Schrimpf, Tobias Steinke. Langzeitarchivierungs-Policy der Deutschen Nationalbibliothek, Version 1.1, 2014, urn:nbn:de:101-2014070801.

Publikationen mit den ihnen eigenen Qualitätsreferenzen“⁴ und zwischen Webseiten. Im ersten Fall werden die Netzpublikationen – analog zur Ablieferung von Printerzeugnissen, Hörbüchern oder CD-ROMs – mithilfe geeigneter Verfahren möglichst vollständig gesammelt. Bei Webseiten ist das aufgrund der großen Qualitätsunterschiede und der großen Menge von nicht-gewerblichen und nicht-institutionellen Netzauftritten nicht möglich. Die Begründung zum DNBG weist ausdrücklich auf diesen Umstand hin: „Die Menge ständig neu erscheinender digitaler Publikationen im Internet (...) schließt konventionelle Methoden und Mittel für die Sammlung, Archivierung und Bereitstellung aus. Eine regelmäßige, stetige Suche, Identifizierung, intellektuelle Erschließung oder eine Sichtung nach Bedeutung ist insbesondere aus finanziellen Gründen ausgeschlossen.“⁵ Für Webseiten wird deshalb auch die „stichprobenweise Sammlung von Internetquellen über Harvesting-Verfahren“ festgelegt.⁶ Dieses Verfahren unterscheidet sich deswegen auch vom Ablieferungsverfahren, das die DNB für Netzpublikationen anbietet.

Einige Ausnahmen von der Sammelpflicht ergeben sich aus § 9 der PflAV. Zu den Ausnahmen gehören unter anderem privaten Zwecken dienende Websites, zeitlich befristete unkörperliche Vorab- und Demonstrationsversionen zu sammelpflichtigen Medienwerken, selbständig veröffentlichte Betriebssysteme, E-Mail-Newsletter ohne Webarchiv und selbständig veröffentlichte Primär-, Forschungs- und Rohdaten. Diese Publikationstypen müssen lediglich gesammelt werden, wenn sie „in physischer oder in elektronischer Form erkennbar zu den ablieferungspflichtigen Netzpublikationen gehören (...). Dies gilt insbesondere für nicht marktübliche Hilfsmittel, die eine Bereitstellung und Benutzung der Netzpublikationen erst ermöglichen und bei den Ablieferungspflichtigen erschienen sind. Sie sind zusammen mit den Netzpublikationen abzuliefern oder zur elektronischen Abholung bereitzustellen“ (§7, PflAV). Bei unterschiedlichen technologischen, sonst aber inhaltsgleichen Ausführungen von Netzpublikationen genügt zudem die Ablieferung bzw. Sammlung einer Version. Näher ausgeführt werden die Auswahlkriterien in den Sammelrichtlinien. Die Sammelrichtlinien haben Handreichungscharakter für die Mitarbeiterinnen und Mitarbeiter und enthalten Anweisungen, wie beispielsweise: „Zu sammeln sind jedoch 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.“⁷

⁴ Begründung (des Gesetzes über die Deutsche Nationalbibliothek),

http://www.dnb.de/SharedDocs/Downloads/DE/DNB/wir/dnbGesetzBegrueundung.pdf?__blob=publicationFile (zuletzt aufgerufen am 13.11.2017).

⁵ ebd.

⁶ ebd.

⁷ Sammelrichtlinien (der Deutschen Nationalbibliothek), Stand: 1. Mai 2014, urn:nbn:de:101-2012022707, S. 57.

Netzpublikationen, die in dieses Sammelgebiet fallen, gilt es nun gemäß § 2, Absatz 1 DNBG „auf Dauer zu sichern und für die Allgemeinheit nutzbar zu machen“, was nichts Anderes bedeutet, als dass die digitale Langzeitarchivierung (LZA) unmittelbar aus dem gesetzlichen Auftrag der DNB hervorgeht. Sie ist auch deswegen eine große Herausforderung, weil die Anzahl der zu archivierenden Netzpublikationen von 3,5 Millionen (2016) bis ins Jahr 2021 nach heutiger Schätzung auf 26 Millionen steigen wird.

Ablieferungsverfahren für Netzpublikationen

Seit der Veröffentlichung des DNBG im Jahre 2006 hat die Deutsche Nationalbibliothek eine LZA-Strategie entwickelt und umgesetzt. Um Netzpublikationen adäquat zu sammeln, musste zunächst ein geeignetes Ablieferungsverfahren entwickelt werden. Die unterschiedlichen organisatorischen und technischen Voraussetzungen der Ablieferer sowie die unterschiedlichen Ablieferungsmengen machten dabei sehr schnell deutlich, dass eine Schnittstelle für die Ablieferung von Netzpublikationen nicht ausreicht, weswegen die DNB nunmehr ihren Ablieferern ein Verfahren mit drei Schnittstellen zur Verfügung stellt: Die OAI-Harvesting-Schnittstelle, Ablieferung via SFTP oder WebDav und die Ablieferung per Webformular. Voraussetzung für die Teilnahme am Ablieferungsverfahren ist, dass der Ablieferer bei der DNB registriert und freigeschaltet wurde.⁸

OAI-Harvesting ist auf die regelmäßige Ablieferung großer Mengen angelegt. Die Ablieferung erfolgt dabei über das OAI Protocol for Metadata Harvesting (OAI-PMH), das den genauen Ablauf des Harvesting-Vorgangs beschreibt und das beide Seiten implementieren müssen. Über diese Schnittstelle sammelt die DNB Netzpublikationen in regelmäßigen Abständen direkt vom Repository des Ablieferers (etwa einem Verlag oder einer Universität) ein. Die auf der Basis des OAI-Protokolls angestoßenen Prozesse arbeiten eine Liste mit Metadaten ab, die der Ablieferer der Schnittstelle zur Verfügung stellt. Darin ist auch die Transfer-URL der Publikation verzeichnet, worüber das zu archivierende Objekt abgeholt werden kann.

Ablieferer haben alternativ die Möglichkeit Netzpublikation mittels SFTP oder WebDav auf einem Server der DNB abzulegen. Dort abgelieferte Dokumente werden über automatisierte Prozesse kontinuierlich in den Pflichtablieferungsworkflow eingespeist. Abgeliefert werden einzelne Transferpakete, die gezippt als Container jeweils eine Netzpublikation mit dem dazugehörigen Metadatensatz enthalten. Auch auf diesem Weg können große Mengen abgeliefert werden.

Die Weiterbearbeitung der Dokumente nach der Ablieferung laufen im Pflichtablieferungs-Workflow unabhängig von der Form der Ablieferung

⁸ Eine ausführliche Beschreibung der Ablieferungsverfahren findet sich in Cornelia Diebel. Netzpublikationen – Sammlung, Archivierung und Bereitstellung in der Deutschen Nationalbibliothek, in: Dialog mit Bibliotheken Bd. 27, 2015, Nr. 1: 24-30, urn:nbn:de:101-2015100136.

vollautomatisiert ab. Es gibt jedoch auch die Möglichkeit geringe Datenmengen manuell abzuliefern. Ein solches Verfahren eignet sich für Ablieferer, die nur gelegentlich einige wenige Publikationen an die DNB abgeben. Im Gegensatz zu den beiden anderen Methoden kann man dabei alle relevanten Daten in ein Webformular eintragen und per Klick abliefern. Die eigentliche Netzpublikation wird entweder durch Angabe einer URL referenziert oder direkt hochgeladen und in den Pflichtablieferungs-Workflow der DNB eingespeist, der von da an ebenfalls automatisiert abläuft.

Die abgelieferten Netzpublikationen werden zunächst auf ihre Vollständigkeit geprüft und danach im sog. Pre-Ingest auf die Langzeitarchivierung vorbereitet. Der Pre-Ingest ist Teil des Import-Prozesses, bei dem die Netzpublikationen automatisch in die Systeme der DNB aufgenommen werden. Dieser Vorgang umfasst mehrere Schritte und Prüfroutinen, von denen die Dateiformat-Prüfung, der Metadaten-Import- und Aufbereitung sowie die Vergabe der URN nachfolgend etwas eingehender beschrieben werden.

Dateiformate und Ingest-Level

Zur Sicherung der Authentizität und langfristigen Nutzbarkeit der abgelieferten Dateiformate führt die DNB beim Import für jede Netzpublikation eine Qualitätsprüfung durch. Dabei arbeitet sie gemäß ihrer Dateiformat-Policy mit sogenannten Ingest-Leveln.⁹ Der Ingest-Level ist das Ergebnis eines mehrstufigen aufeinander aufbauenden Prüfverfahrens, das während des Importprozesses einer Netzpublikation automatisiert abläuft. Dabei werden fünf verschiedene Qualitätskriterien nacheinander abgeprüft.

1. Dateintegrität (Checksummenprüfung; Level 0)
2. Identifikation des Dateiformats (Level 1)
3. Beschränkungsfreiheit (z.B. keine Verschlüsselung durch Passwörter; Level 2)
4. Extraktion formatspezifischer technischer Metadaten möglich (Level 3)
5. Format kann validiert werden (Level 4)

Je mehr Kriterien eingehalten werden, desto höher der Ingest-Level. Die DNB verwendet grundsätzlich eine Default-Format-Policy. Dabei muss bei der Ablieferung von Netzpublikationen mindestens Ingest-Level 2 erreicht werden. Nur in Ausnahmefällen werden Abweichungen davon mit den Ablieferern abgesprochen. Wird Level 2 oder das individuell vereinbarte Level nicht erreicht, erhält der Ablieferer eine Fehlermeldung mit der Bitte, das Dateiformat der Publikation gemäß des vereinbarten Levels aufzubereiten.

⁹ Vgl. Spezifikation der Dateiformat-Policy für die Sammlung von Netzpublikationen der Deutschen Nationalbibliothek: Erläuterungen zur Handhabung, Version 1.0, Stand: 24.10.2012, urn:nbn:de:101-2012102408.

Metadaten

Bei der Automatisierung ihrer Geschäftsgänge profitierte die DNB von der Herausbildung diverser Metadaten-Standards, die nunmehr für den Import der deskriptiven Metadaten einer Netzpublikation vorgeschrieben werden können. So wurde im Verlagswesen der Metadatenstandard ONIX entwickelt, um über dieses Datenformat Informationen zum Buch auf Webseiten, Daten für die Meldung an das VLB und ggf. auch Daten für die Presse bzw. die Verkaufskataloge bereitzustellen. Auch bei der Pflichtablieferung leistet ONIX gute Dienste. Für die Ablieferung von Hochschulprüfungsarbeiten entwickelte die DNB selbst das Format XMetaDissPlus (XMDP). Weitere Formate, die bei der Pflichtablieferung von Netzpublikationen akzeptiert werden, sind MARCXML und für Zeitschriftenartikel NISO JATS.

Daten, die für einen Nachweis im Katalog der DNB minimal erforderlich sind - und zwar unabhängig von konkreten Standards - werden in sogenannten Kernsets für verschiedene Publikationstypen (Monographien, Zeitschriften etc.) dokumentiert. Anhand dieser Kernsets bauen die Ablieferer den Metadatensatz der jeweiligen Netzpublikation auf. Darüber hinaus können auch fakultative Angaben gemacht werden, die ebenfalls in den Kernsets definiert sind. Bei der Ablieferung wird dann dieser Datensatz auf Vollständigkeit geprüft, d.h. die im Kernset vorgegebenen obligatorischen Metadaten müssen vollständig vorliegen.¹⁰ Während des Importprozesses werden diese Daten in das von der DNB genutzte Bibliotheksformat PICA umgewandelt und mit Standardwerten und anwendungsspezifischen Feldern angereichert. Als PICA+-Datensatz werden die Metadaten in das Katalogsystem (s. Abbildung 1 weiter unten) der DNB übernommen. Zudem findet eine Dublettenprüfung statt.

Von den deskriptiven Metadaten müssen die technischen Metadaten unterschieden werden, die für das Preservation Planning relevant sind. Diese Daten werden nicht von Verlagen oder Hochschulen bereitgestellt, sondern während des Importprozesses von der DNB automatisiert generiert. Sie umfassen charakteristische Eigenschaften wie Dateiformat, Formatversion oder Dateigröße. Die DNB setzt bei der Datenerhebung eine Reihe etablierter Tools ein, die in FITS (File Information Tool Set) zusammengefasst sind.¹¹ Die erhobenen technischen Metadaten werden im XML-basierten Format LMER¹² erfasst, das auch in METS integrierbar ist, und in der LZA-Verwaltungsdatenbank (s. Abbildung 1) abgelegt. Wenn Migrationsschritte

¹⁰ Vgl. Lieferung von Metadaten für monografische Netzpublikationen an die Deutsche Nationalbibliothek: Metadaten-Kernset Monografien, Hörbücher und Hochschulprüfungsarbeiten. Version 2.0, Stand: 15. August 2014, urn:nbn:de:101-2014071100.

¹¹ FITS ist ein Toolset, das diverse Metadaten-Tools bündelt. So zum Beispiel JHOVE 1 für die Validierung von Formaten. Vgl. hierzu auch die Homepage von FITS, <https://projects.iq.harvard.edu/fits> (zuletzt aufgerufen am 13.11.2017).

¹² Vgl. LMER : Langzeitarchivierungsmetadaten für elektronische Ressourcen. Version 1.2, Stand: 7. April 2005, urn:nbn:de:1111-2005041102.

anstehen, werden diese Daten analysiert, die zu migrierenden Objekte identifiziert und entsprechend verarbeitet.

Persistente Identifizierung

Alle von der Deutschen Nationalbibliothek archivierten Netzpublikationen erhalten eine URN (Uniform Resource Name) aus dem Namensraum "urn:nbn:de". Im Gegensatz zu URLs bieten URNs eine eindeutige und dauerhafte Identifikation von Objekten, unabhängig vom Ort der Speicherung. Ändert sich der Speicherort der Publikation, z.B. durch eine Serverumstellung, kann die Zugriffsadresse (URL), die zu einer URN hinterlegt ist, im Resolving-System der DNB durch den Ablieferer korrigiert werden. In der Regel sind aber mehrere URLs pro URN registriert, sodass beim Ausfall einer Ressource die Netzpublikation weiterhin zugreifbar ist. Dadurch behält die URN ihre Gültigkeit und verweist nach wie vor auf die entsprechende Publikation. Das Zitieren einer URN ist somit ein sicherer Weg, auch langfristig auf ein digitales Objekt zu verweisen. In der Regel generiert die abliefernde Institution die URN für eine Netzpublikation, wobei dieser Vorgang durch die Vergabe von Unternehmensräumen und Identifikatoren für die jeweilige Publikation eindeutig festgelegt ist. Beim Import wird geprüft, ob eine URN vorhanden ist; ansonsten generiert der Import-Prozess sie durch die Vergabe von Standardparametern automatisch.

Web-Harvesting

Wie weiter oben schon erörtert, kann bei Webseiten nicht von einer Ablieferung, sondern eher von einer „stichprobenweisen Sammlung“ gesprochen werden. Im Gegensatz zur schon beschriebenen Ablieferung durch OAI-Harvesting gibt es beim Web-Harvesting-Verfahren keinen Ablieferer, der eine Liste zusammenstellt, anhand der dann das Harvesting durchgeführt wird. Vielmehr wählen Mitarbeiterinnen und Mitarbeiter der Deutschen Nationalbibliothek die zu sammelnden Webseiten nach festgelegten inhaltlichen und formalen Kriterien aus. Die DNB sammelt somit selektiv einen vorab festgelegten Ausschnitt des Web-Angebots. Der Fokus liegt derzeit auf Webseiten ausgewählter Institutionen (z.B. Bundesbehörden, Interessenverbänden, Kultureinrichtungen usw.) sowie auf Webseiten zu bestimmten Ereignissen (etwa Bundestagswahlen). Das Sammelspektrum soll aber zukünftig erweitert werden.

Das Harvesting führt ein externer Dienstleister über ein automatisches Verfahren anhand des von der DNB vorgegebenen Sammelspektrums durch. Dabei wird grundsätzlich von einer Start-URL ausgegangen, über die eine zum Sammelspektrum gehörende Webseite sowie alle verlinkten Inhalte abgespeichert werden, bis ein Link zu einer anderen Domain führt. Dieses Verfahren ermöglicht eine Momentaufnahme der Webseite, die man Snapshot oder Zeitschnitt nennt. Dieser Zeitschnitt wird

gemäß der LZA-Policy als abgeschlossene Publikation betrachtet und archiviert. Um der dynamischen Publikationsweise im Web gerecht zu werden, werden in bestimmten Abständen weitere Zeitschnitte durchgeführt, um die Entwicklung des jeweiligen Webauftrittes anhand der archivierten Zeitschnitte nachvollziehen zu können. Die Sammelfrequenz kann für jede Webseite individuell eingestellt werden, je nach Änderungshäufigkeit bzw. inhaltlicher Relevanz. Im Moment erfolgt ein Zeitschnitt in der Regel halbjährlich. Ergänzend zu diesem Vorgehen wurde 2014 auch ein Crawl der Top-Level-Domain .de durch die Internet Memory Foundation im Auftrag der DNB durchgeführt.

Übertragung ins Langzeitarchiv und Datenhaltung

Nach Abschluss des Imports werden aus den abgelieferten und mit Metadaten versehenen Netzpublikationen Archivobjekte im offen definierten Paketformat Universelles Objektformat (UOF)¹³ generiert. Dabei 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 OAIS-konforme Archivsystem (s. Abbildung 1) – eingespielt, in dem der gespeicherte Bestand 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.

Zudem sollen in absehbarer Zeit auch Emulationen möglich sein. Dazu werden die im DFG-Projekt EMiL (Emulation of Multimedia objects in Libraries) entwickelten Tools in die Bereitstellungsarchitektur der DNB integriert. Auf diese Weise wird es möglich sein, historische Betriebssysteme auf den aktuell eingesetzten Plattformen nachzubilden, um insbesondere komplexe digitale Publikationen authentisch erhalten zu können. Objekte, die bei alleiniger Anwendung der Migrationsstrategie aufgrund proprietärer Formate und Systemanforderungen oder aufgrund ihrer Komplexität

¹³ UOF wurde im Rahmen des Projekts kopal auf Basis von LMER und METS entwickelt und spezifiziert. Vgl. Tobias Steinke. Universelles Objektformat: Ein Archiv- und Austauschformat für digitale Objekte. 2006, http://kopal.langzeitarchivierung.de/downloads/kopal_Universelles_Objektformat.pdf (zuletzt aufgerufen am 13.11.2017).

nicht mehr benutzbar wären, können so dauerhaft im Lesesaal aufgerufen und in ihren originären Systemumgebungen verwendet werden.¹⁴

Die technische Infrastruktur

Die DNB hat in den vergangenen Jahrzehnten eine umfangreiche technische Infrastruktur aufgebaut. Diese bildet die Grundlage für die weiter oben schon beschriebenen Ablieferungs- und Importprozesse, für die Langzeitarchivierung und für die Bereitstellung der archivierten Objekte im Rahmen der Benutzung.

Die technische Basis der Langzeitarchivierung selbst bildete bis 2017 das DIAS-Archivsystem, das im kopal-Projekt entwickelt wurde. Gefördert vom Bundesministerium für Bildung und Forschung (BMBF) baute 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 auf. 2017 wurde DIAS von der GWDG durch koala (Kooperative Langzeitarchivierung) ersetzt. Das Archivsystem koala ist ebenso wie DIAS OAIS-konform; es besteht aus Open-Source-Komponenten und garantiert somit die Anschlussfähigkeit der Langzeitarchivierung an neueste Entwicklungen. Über das Bereitstellungssystem kann dann der im DNB-Portal angemeldete Benutzer auf die Metadaten aus dem Katalogsystem und die Objekte zugreifen, die dazu – neben dem eigentlichen Archivsystem – im DNB-eigenen Repository aufbewahrt werden. Abbildung 1 zeigt eine schematische Darstellung der gesamten technischen Infrastruktur der DNB.

¹⁴ Vgl. Nathalie Lubetzki, Tobias Steinke. EMIL - Emulation von Kulturgut, in : Dialog mit Bibliotheken Bd. 29, 2017, Nr. 1: 20-24, urn:nbn:de:101-20170309140.

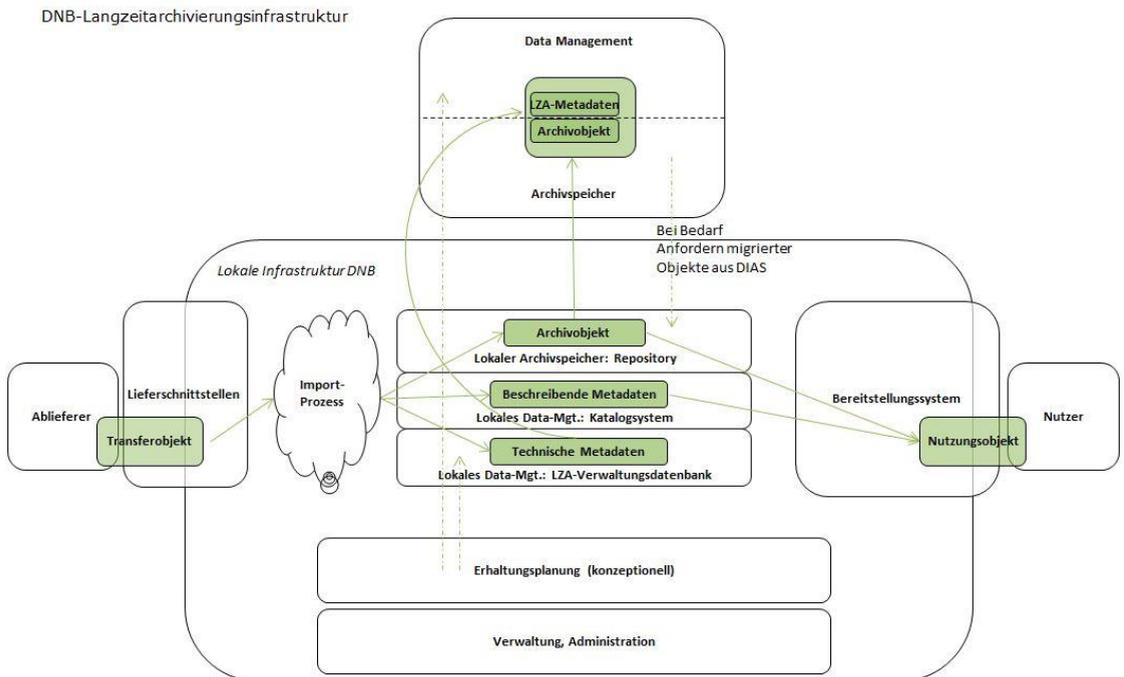


Abbildung 1: Die technische Infrastruktur der DNB für die Ablieferung und die Langzeitarchivierung digitaler Objekte

Langzeitarchivierungs-Policy der DNB

Gut zusammenfassen lässt sich die Langzeitarchivierungs-Strategie der DNB in den 13 Prinzipien der am 08.07.2014 veröffentlichten LZA-Policy:¹⁵

1. Gleichbehandlung der zu archivierenden Medien unabhängig von Inhalt und technischer Komplexität
2. Nutzung automatisierter Verfahren bei Ablieferung und Verarbeitung digitaler Medien
3. Sorgfalt bei der Gestaltung der Ablieferungs- und Verarbeitungsprozesse und Minimierung dabei auftretender Fehlerquellen
4. Sicherung der Integrität der archivierten Medien
5. Erhaltung der Authentizität der archivierten Medien
6. Dauerhafte Zugänglichkeit zu den archivierten Medien
7. Abgeschlossenheit; Dynamische Objekte (wie Webseiten, Forschungsdaten) werden bei der Ablieferung an die DNB in dem zu diesem Zeitpunkt bestehenden Zustand als abgeschlossen betrachtet.

¹⁵ Schrimpf, Steinke (2014). Langzeitarchivierungs-Policy der Deutschen Nationalbibliothek

8. Eindeutige Identifizierbarkeit der archivierten Medien durch URN
9. Aktives und vorausschauendes Handeln bei der Langzeitarchivierung
10. Erhaltungsmaßnahmen werden auf Bestandsgruppen-Ebene geplant und durchgeführt
11. Vertrauenswürdigkeit des Langzeitarchivs durch regelmäßige Zertifizierungen
12. Standardkonformität; Verwendung von und Mitarbeit bei der Entwicklung von LZA-Standards (z.B. OAIS, DIN 31644, nestor-Leitfäden)
13. Kontinuierliche Weiterentwicklung der gesamten Langzeitarchivierungsinfrastruktur

Als besonders wichtige Prinzipien sollen die Automatisierung und die Zertifizierung hervorgehoben werden. Die vollständige Automatisierung der Prozesse von der Ablieferung bis zur Einspeisung einer Publikation in das Langzeitarchiv ist mittlerweile komplett umgesetzt. Manuelle Eingriffe sind somit nur noch in Ausnahmefällen notwendig. Außerdem erwirbt die DNB Zertifizierungen für ihre Systeme und Prozesse. So wurde das Archivsystem 2015 nach den Richtlinien des Data Seal of Approval als vertrauenswürdige Langzeitarchiv zertifiziert. 2016 erhielt die DNB für den Bereich der Netzpublikationen das nestor-Siegel. Die Zertifizierungen werden in regelmäßigen Abständen überprüft und erneuert.

Ausblick

Welche Herausforderungen und Aufgaben stellen sich für die DNB jetzt und in der nahen Zukunft? Besonders hohen Stellenwert genießt derzeit die organisatorische und technische Anpassung der Langzeitarchivierung auf die ständig steigende Zahl unkörperlicher Medienwerke. Dies ist nötig, um in den kommenden Jahren die prognostizierten Ablieferungszahlen von 2,5 bis 3 Millionen jährlich zu bewältigen. Hinzu kommen auch noch neue Sammelobjekte wie z.B. Netzpublikationen der Musik und die Ergebnisse der Digitalisierungsprojekte. Derzeit werden unter anderem vergiffene Bücher, Adressbücher und AV-Medien digitalisiert. Bis 2020 soll dieser Innovationszyklus abgeschlossen sein.

Außerdem sollen an dieser Stelle noch weitere Projekte genannt werden, die aktuelle Entwicklungen adressieren. Weiter oben wurde bereits auf das Projekt EMiL hingewiesen, mit dem die DNB ihre Langzeitarchivierungsstrategie um die Emulation erweitert. Durch die Implementierung der Ergebnisse dieses Projektes kann die dauerhafte Zugänglichkeit komplexer Objekte (z. B. Multimediaobjekte, digitale Kunstwerke, wissenschaftliche Simulationen) weitgehend unabhängig von der Verfügbarkeit veraltender Plattformen und Betriebssysteme sichergestellt werden.

Mit Forschungsdaten beschäftigt sich das Projekt eDissPlus (Elektronische Dissertationen Plus). Aufgrund der gesetzlichen Vorgaben sammelt die DNB zwar

keine selbstständig publizierten Forschungsdaten, wohl aber lässt sich aus der Pflichtabgabeverordnung die Verantwortung für Daten ableiten, die als Supplement einer Publikation beigegeben wurden. Das sind zum Beispiel Forschungsdaten, die im Rahmen eines Dissertationsprojektes entstehen oder die die Promovenden für ihre Arbeit nachnutzen. In eDissPlus entwickelt die DNB zusammen mit der Humboldt-Universität einen Geschäftsgang, der die Ablieferung, Langzeitarchivierung und Benutzung dissertationsbezogener Daten ermöglicht. Dabei wird auch die persistente Identifizierung dieser Daten geklärt und die Frage, wie Bezüge zwischen Forschungsdaten und Dissertation in den Metadaten abgebildet und in der Benutzung vermittelt werden können.¹⁶

Neben dem Aufbau und Betrieb des eigenen Langzeitarchivs sieht die DNB gerade bei kleineren Kulturerbeeinrichtungen einen Bedarf, Langzeitarchivierung als Dienstleistung einkaufen und nutzen zu können. Wegen des hohen Aufwands ist es für solche Einrichtungen meist nicht wirtschaftlich, eigene Langzeitarchivierungsprozesse und Infrastrukturen aufzubauen und das notwendige Know-How zum Betrieb vorzuhalten. Durch einen hohen Grad der Standardisierung lassen sich erhebliche Synergieeffekte bei einer gemeinsamen Nutzung von Langzeitarchivierungsinfrastrukturen erzielen. Die DNB möchte ein entsprechendes Dienstleistungsportfolio entwickeln und Kulturerbeeinrichtungen anbieten.

Am Beispiel der Netzpublikationen schilderte dieser Beitrag den aktuellen Stand der Langzeitarchivierung in der DNB. Seit 2006 wurden ausdifferenzierte Geschäftsgänge für die Sammlung und Archivierung nichtkörperlicher Medienwerke geschaffen. Die Mitarbeiterinnen und Mitarbeiter der DNB konnten dabei auch die Erfahrung machen, dass sich mit der Weiterentwicklung der elektronischen Medien auch die Anforderungen an die Ablieferung und Archivierung von Netzpublikationen und anderen digitalen Objekten verändern. Zudem werden auch neue Medientypen wie z.B. Forschungsdaten vom Sammelauftrag erfasst. Die abschließend geschilderten Aktivitäten zeigen beispielhaft, wie die DNB auf diese Anforderungen reagiert und damit ihrem Auftrag als Archivbibliothek und nationalbibliografischem Zentrum nachkommt.

¹⁶ Vgl. Dirk Weisbrod. Forschungsdaten in Dissertationen, in: Dialog mit Bibliotheken Bd. 28, 2016, Nr. 2: 29-30, urn:nbn:de:101-20161006215.