



Umsetzung der Pflichtablieferung für Netzpublikationen in Deutschland

Renate Gömpel
&

Dr. Lars G. Svensson
Deutsche Nationalbibliothek
Frankfurt, Germany

Persistent Identifier: urn:nbn:de:101-2011061609

Meeting:

193 — e-Legal deposit: from legislation to implementation; from ingest to access — Bibliography Section with IFLA-CDNL Alliance for Digital Strategies Programme (ICADS), Information Technology, National Libraries and Knowledge Management

Abstract:

Ein neues Gesetz über die Deutsche Nationalbibliothek ist 2006 in Kraft getreten. Die umfangreichste Änderung gegenüber dem vorangegangenen Gesetz bezieht sich auf das elektronische Pflichtexemplarrecht, die Sammlung, Katalogisierung, Indexierung und Archivierung von Ressourcen in unkörperlicher Form.

Seither implementiert die Deutsche Nationalbibliothek einen voll-automatisierten Workflow für die Erwerbung, Katalogisierung und Archivierung aller Arten von Online-Materialien. Durch die Nutzung dieses Workflows können wir E-Books, Online-Hochschulschriften, E-Journals und digitale Zeitungen bearbeiten und wir arbeiten daran, Websites zu archivieren.

Die Erstellung des Workflows umfasste die Behandlung von Metadatenformaten aus verschiedenen Quellen, wie z. B. Verlage, akademische Gesellschaften und Einzelpersonen, aber auch von verschiedenen Dateiformaten, die archiviert werden müssen. Auch die Beachtung des Urheberrechts gehörte dazu.

Der Workflow vereinfacht die Ablieferung von Netzpublikationen soweit wie möglich. Aktuell haben wir drei Übertragungs-Schnittstellen:

- über ein Webformular (für kleine Mengen von Publikationen)*
- über OAI PMH (z. B. für Verlage und Universitäten)*
- über ftp oder WebDAV in einen Hotfolder (meist von kommerziellen Verlagen genutzt).*

In allen drei Fällen werden Metadaten vom Verfasser oder vom Verlag mitgeliefert und ohne intellektuelles Eingreifen in den Katalog übertragen. Sobald die Ressource in das Repository übertragen und archiviert ist, ist der Titel im Katalog zu sehen und die Publikation kann im Lesesaal gelesen werden. Alle neu übermittelten Netzpublikationen werden in einer speziellen Reihe der Deutschen Nationalbibliografie verzeichnet, der Reihe O.

Die Langzeitarchivierung wird garantiert, aber bisher lassen sich nicht alle Arten von Dateiformaten abspeichern. Die Deutsche Nationalbibliothek arbeitet daran, das Spektrum der verarbeitbaren Metadaten und Dateiformate zu erweitern, um den Herausforderungen des elektronischen Zeitalters gerecht zu werden. Um sicherzustellen, dass unsere gespeicherten digitalen Ressourcen verlässlich zitiert werden können und zugänglich sind, vergeben wir persistente Identifikatoren für alle Netzpublikationen, die als Pflichtexemplare in die Bibliothek kommen. Der Einsatz der persistenten Identifizierung stellt nicht nur sicher, dass die Ressource identifizierbar und zugänglich bleibt, auch wenn sie ihren Speicherort ändert oder migriert wird, sondern sie ist auch wichtig im weiteren Kontext der Nationalbibliografien und für Aggregatoren wie z. B. die Europeana oder die Deutsche Digitale Bibliothek, die Daten aus verschiedenen Quellen zusammenstellen.

Der Vortrag stellt den Workflow der Deutschen Nationalbibliothek für Ressourcen in unkörperlicher Form sowie die Nutzung von persistenten Identifikatoren für die dauerhafte Zugänglichkeit dar.

Einleitung

Die Deutsche Nationalbibliothek (DNB) hat die Aufgabe, alle deutschen und deutschsprachigen Medienwerke ab 1913 zu sammeln, zu inventarisieren, zu erschließen und bibliografisch zu verzeichnen sowie auf Dauer zu sichern und der Allgemeinheit nutzbar zu machen. Dieser Auftrag ist seit Inkrafttreten des Gesetzes über die Deutsche Nationalbibliothek (DNBG) vom 22. Juni 2006 auf die Sammlung von Medienwerken in „unkörperlicher Form“ erweitert, für die sich der Begriff Netzpublikationen etabliert hat.

Der Auftrag ist nicht nur im DNBG genannt, sondern wird in der Pflichtablieferungsverordnung präzisiert. In dieser Rechtsverordnung sind die Bedingungen genannt, wann und wie Medienwerke (einschließlich Netzpublikationen) abzuliefern sind, sowie der Umfang und die Einschränkungen der Ablieferungspflicht.

Nach Inkrafttreten der Pflichtablieferungsverordnung am 17. Oktober 2008 wurden als Arbeitsinstrument zur differenzierten Festlegung der Sammelpflicht auch anhand von Einzelbeispielen die Sammelrichtlinien in einer neuen Fassung veröffentlicht, die in einem separaten Teil auch Aussagen über Netzpublikationen enthält.

Was sind Netzpublikationen?

Unter die Sammelpflicht fallen alle Darstellungen in Schrift, Bild und Ton, die in öffentlichen Netzen zugänglich gemacht werden. Beispiele sind E-Books, E-Journals, Digitalisate, Musikdateien oder auch Webseiten. Die Frage, welche elektronischen Produkte von der

Sammelpflicht erfasst werden, wird in der Neufassung der Pflichtablieferungsverordnung und den Sammelrichtlinien beantwortet (siehe auch <http://files.d-nb.de/pdf/sammelrichtlinien.pdf>).

Welche technischen Möglichkeiten der Ablieferung gibt es?

Wie bei den gedruckten Publikationen sind alle, die Netzpublikationen veröffentlichen, sowohl kommerzielle als auch private Organisationen oder Personen, gesetzlich verpflichtet, ihre Publikationen an die Deutsche Nationalbibliothek abzuliefern. Nicht alle sind mit der Übermittlung von Dateien und Metadaten vertraut; deshalb war es notwendig, einen einfachen Weg zu finden, der einerseits die Ablieferung ohne zu großen Aufwand ermöglicht und andererseits die Deutsche Nationalbibliothek in die Lage versetzt, die große Menge an elektronischen Publikationen weitgehend automatisiert zu verarbeiten.

Aufbau eines automatisierten Workflows – stufenweise Umsetzung

Als Folge des neuen gesetzlichen Auftrags begann die Deutsche Nationalbibliothek mit dem Aufbau eines Workflows für die automatische Ablieferung von Netzpublikationen, deren Anzeige im Katalog und die Archivierung der Dateien.

Dafür wurde ein stufenweises Vorgehen gewählt, was den Vorteil bietet, die Technik mit kleineren Datenmengen im Echtbetrieb testen und verfeinern zu können.

Als erster Schritt ist bei allen eingesetzten Verfahren und Schnittstellen die Registrierung der Ablieferer von Netzpublikationen vorgesehen. Die dabei erfassten Daten (wie z. B. Name, Adresse und E-Mail-Adresse) sind nur für den Ablieferer und die Mitarbeiter der Deutschen Nationalbibliothek sichtbar, die im Bereich Netzpublikationen arbeiten. Bei der Registrierung wird auch der Fall berücksichtigt, dass Dienstleister, die für Verlage die elektronischen Medien vertreiben, im Auftrag abliefern.

Ablieferung über Webformulare

Webformulare waren die erste Möglichkeit zur Ablieferung der verschiedenen Arten von Netzpublikationen. Aus einem ersten Formular für Monografien (E-Books) und einem zweiten Formular für Online-Hochschulschriften, das bereits vor 2006 auf freiwilliger Basis genutzt wurde, wurden die neuen Formulare entwickelt, die eine einfache Übermittlung weiterer Netzpublikationen erlauben: das Formular für monografische Veröffentlichungen kann für E-Books, Online-Hochschulschriften und Noten genutzt werden, wobei es nur wenige gemeinsame Pflichtfelder (wie Titel, Veröffentlichungsdatum, Adresse der Veröffentlichung) gibt, je nach Publikationstyp aber weitere Felder hinzukommen können (z. B. Angaben über die Dissertation bei Online-Hochschulschriften oder spezielle Identifikatoren wie die ISMN bei Noten). Ein weiteres Formular dient zur Übermittlung von Titeln von elektronischen Zeitschriften, ein drittes zur Übermittlung von Zeitschriftenlieferungen. Zwischen diesen Formularen besteht ein Zusammenhang, da ein Titel nur einmal gemeldet wird und der Ablieferer bei der nächsten Anmeldung eine Liste seiner Zeitschriften erhält und durch Anklicken die meisten Felder für eine Zeitschriftenlieferung bereits automatisch ausgefüllt werden.

Die Webformulare sind für kleinere Mengen von Publikationen geeignet, da die Eingabe der Metadaten manuell erfolgt. Es gibt außerdem eine Obergrenze von 50 MB für das Hochladen von Einzelobjekten und von 500 MB für das Abliefern per URL.

Bereitstellung zur Abholung über eine Schnittstelle

Als nächste Stufe wurde eine Schnittstelle eingerichtet, über die Publikationen vom Ablieferer bereitgestellt und von der Deutschen Nationalbibliothek abgeholt werden. Dafür wird ein HTTP-basiertes Harvesting-Protokoll genutzt, das von der Open Archive Initiative entwickelt wurde (OAI-PMH, Open Archive Initiative, Protocol for Metadata Harvesting). Bei der Nutzung von OAI-PMH fordert ein Client oder Harvester durch HTTP-GET-Requests an einen Server oder ein Repository Daten an. Damit werden in einem vollautomatisierten Verfahren, das auf keiner der beiden Seiten eine manuelle Interaktion erforderlich macht, von der DNB Metadaten vom Server des Ablieferers abgeholt. In den Metadaten ist eine Transfer-URL enthalten, über welche die Publikation ebenfalls automatisch abgeholt wird. Die Metadaten werden ohne die Transfer-URL in den Katalog der DNB übertragen und die Dateien in das Repository. Weitere Informationen zur Einrichtung einer OAI-Schnittstelle sind unter http://www.d-nb.de/netzpub/abliefer/pdf/automatisierte_ablieferung.pdf zu finden. Dieses Verfahren läuft nach einer Testphase auf beiden Seiten automatisch ab und eignet sich besonders für größere Mengen von Dateien.

Ablieferung über Hotfolder

Eine weitere Schnittstelle wird seit April 2011 eingesetzt. Hotfolder eignen sich zur Übermittlung von größeren Datenmengen, die vom Ablieferer an diesen überwachten Ordner geschickt werden. Der Ordner wird „hot“ genannt, weil jeder Schritt eines darin ablaufenden Prozesses von einem anderen Prozess überwacht wird. Nach Anlegen eines Ablieferer-Accounts werden die Publikationen zusammen mit den Metadaten in einem Zip-Container abgelegt. Mögliche Übertragungswege für den Container sind FTP (File Transfer Protocol) oder auch WebDAV. Durch ein automatisches Verfahren werden die Metadaten in den Katalog der DNB integriert und die Dateien im Repository archiviert. Der Hotfolder erfordert aktives Bereitstellen der Publikationen sowie der Daten durch den Ablieferer, wurde aber von den Verlagen wegen der bekannten Transfermöglichkeiten (wie FTP) gewünscht.

Bei allen drei Verfahren werden sowohl die Metadaten oder Titeldaten als auch das elektronische Objekt selbst gesammelt.

Handhabung von Dateiformaten

Netzpublikationen werden in dem Dateiformat gesammelt, in dem sie veröffentlicht wurden. Dabei ist zu beachten, dass eine transferfähige Netzpublikation eine selbständige logische Einheit bilden muss, die aus ihrer Umgebung herausgelöst werden kann und nicht auf eine Serveranbindung im Hintergrund angewiesen sein darf, über die wesentliche Inhaltsbestandteile dynamisch und ad hoc zum Zeitpunkt der Interaktion mit dem Nutzer aus einem Datenhaltungssystem abgerufen werden müssen.

Derzeit können neben PDF (PDF/A und alle anderen PDF-Typen) das Format EPUB sowie Dokumente in HTML automatisch archiviert werden.

Alle Dateien müssen unverschlüsselt übermittelt werden, um die dauerhafte Benutzbarkeit der Dokumente mit möglichst wenig Aufwand zu gewährleisten.

Handhabung von Metadatenformaten

Bei der Ablieferung über ein Webformular werden die Metadaten vom Ablieferer selbst eingegeben; dabei sind nur die Vorgaben zu beachten, die im Formular erläutert werden, aber es muss kein Standard eingehalten werden. Die Metadaten aus den ausgefüllten Feldern werden über ein Mapping in die entsprechenden Felder des internen Katalogformats übertragen.

Bei den beiden anderen Ablieferwegen müssen die Metadaten für den automatischen Workflow nach einem definierten und vorher vereinbarten Standard abgeliefert werden. Die derzeit akzeptierten Formate sind ONIX for Books, MARC-XML oder XMetaDissPlus; weitere Formate sollen folgen.

Um die Ablieferung zu erleichtern, wurden in einem Metadatenkernset die minimalen Anforderungen an Metadatenelemente definiert. Einzelheiten sind unter http://www.d-nb.de/netzpub/abliefer/pdf/metadaten_kernset_definitionen.pdf zu finden.

Was geschieht mit den Netzpublikationen?

Automatische Erschließung

Die Metadaten werden weitgehend unverändert in den Katalog der DNB übernommen. Beim Einspielen werden keine Normdatenverknüpfungen gemacht, und nur einige Pflichtfelder werden überprüft. Der Inhalt der Felder wird dabei jedoch nicht geprüft und kann aufgrund der Menge der Datensätze nicht manuell bearbeitet werden. In einem DNB-Projekt werden deshalb Szenarien entwickelt, die durch automatische Erschließung den Informationsgehalt der Datensätze verbessern:

- Die gemeldeten Personennamen (wie z. B. Verfasser, Herausgeber, Übersetzer) werden nachträglich mit der Personennormdatei abgeglichen und mit Normdatensätzen verknüpft.
- Die neuen Datensätze für Netzpublikationen werden nachträglich auf parallele Druckausgaben geprüft und – falls vorhanden – mit ihnen verknüpft, wobei Informationen aus den intellektuell erschlossenen Druckausgaben in die Datensätze der Netzpublikationen übertragen werden.
- Netzpublikationen werden automatisch mit Dewey-Sachgruppen und Schlagwörtern indexiert, insbesondere in den Fällen, in denen es keine parallelen Druckausgaben gibt.

Als Teil eines weiteren Projekts wird Suchmaschinen-Technologie eingesetzt, um die Katalogsuche und die Retrievalprozesse insbesondere für Netzpublikationen zu verbessern.

Bereitstellung der Metadaten im Katalog sowie der Publikationen im Lesesaal

Die Metadaten der abgelieferten und archivierten Netzpublikationen sind im Katalog der Deutschen Nationalbibliothek frei sichtbar, die Rechteverwaltung für die digitalen Objekte ist hingegen komplexer. Bei der Ablieferung eines Objekts kann der Ablieferer angeben, welche Verbreitungsrechte er der DNB einräumt. Die Spannweite der Rechte reicht vom Zugriff durch den Endnutzer nur im Lesesaal der Bibliothek über den Zugang über das Internet für registrierte Benutzer, so dass sie von zuhause aus auf die Publikation zugreifen können, bis hin zum weltweit unbeschränkten Zugang für alle Benutzer. Kommerzielle Verlage bieten normalerweise kostenpflichtigen Zugang zu einer Publikation über ihre Website an und beschränken die Nutzung auf den Zugriff vor Ort in der Bibliothek. In den Fällen, in denen nur im Lesesaal der Bibliothek auf die Publikation zugegriffen werden kann, können die Endnutzer aufgrund der urheberrechtlichen Beschränkungen keine digitalen Kopien der Objekte machen und sie nicht per E-Mail versenden. Informationen zu den rechtlichen Grundlagen sind unter <http://bundesrecht.juris.de/dnbg/index.html> und <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32001L0029:DE:HTML> zu finden.

Deutsche Nationalbibliografie – die Reihe O

Neben dem Katalogeintrag der Metadaten, der unmittelbar nach der Ablieferung eines Objekts erfolgt, werden alle eingehenden Netzpublikationen in der Reihe O (Netzpublikationen) der Deutschen Nationalbibliografie angezeigt. Einmal monatlich wird eine Zusammenstellung der neu eingegangenen Titel erstellt; diese Datei kann von den Nutzern kostenlos abgerufen werden. Die Datensätze liegen in den Austauschformaten MAB und MARC 21 vor. Im Gegensatz zu den anderen Reihen der Nationalbibliografie wird diese Datei nicht als PDF angeboten.

Zitierbarkeit und Archivierung

Alle digitalen Publikationen, die an die Deutsche Nationalbibliothek abgeliefert werden, werden nicht nur in der Nationalbibliografie verzeichnet, sondern sie erhalten auch einen sogenannten persistenten Identifikator (PI). Ein PI ist ein eindeutiger Identifikator, der weltweit zur Identifizierung adressierbarer Objekte wie z. B. Dokumente, Bilder, Tonaufnahmen, Animationen oder Metadatenbeschreibungen eingesetzt wird. Bei der Verwendung eines PI wird ein als Resolver bezeichnetes Verzeichnis zwischen den Namen und die Adresse des digitalen Objekts geschaltet. Die Aufgabe des Resolvers ist, die Zuordnung einer Verknüpfung zwischen dem Namen und der Adresse sicherzustellen. Ändern sich die Adressen, so werden diese im Resolver aktualisiert, so dass der Resolver stets auf eine gültige Adresse verweist. So bleiben die Referenzen stabil und der Aufwand ist überschaubar. Die Trennung der Identifikation der Objekte von deren Speicherort durch eine eindeutige Zeichenkette ist ein Grundprinzip von persistenten Identifikatoren.

Im Rahmen des Projektes Epicur (2002-2005) führte die Deutsche Nationalbibliothek eine PI-Infrastruktur ein, die als Identifikatoren Uniform Resource Names (URN) aus dem Namensraum NBN (National Bibliography Number) verwendet – einem Namensraum der speziell für die Identifikation bibliografischer Ressourcen konzipiert ist. Um eine dezentrale Verwaltung zu ermöglichen, ist urn:nbn geografisch gegliedert, und die Deutsche Nationalbibliothek ist für Deutschland und somit für den Unternamensraum urn:nbn:de zuständig.

Für die Umwandlung der Identifikatoren in Zugriffsadressen betreibt die Deutsche Nationalbibliothek einen Resolver unter <http://nbn-resolving.org/>. Der Resolver löst nicht nur Namen aus dem Deutschen Namensraum urn:nbn:de auf, sondern auch die aus den entsprechenden schweizerischen und österreichischen Namensräumen (urn:nbn:ch bzw. urn:nbn:at). Außerdem bietet er die Möglichkeit, Anfragen nach persistenten Identifikatoren, die er nicht selber verwaltet, weiterzuleiten. So leitet der Resolver auch Anfragen nach urn:nbn aus der Tschechischen Republik, aus Finnland, aus Ungarn, den Niederlanden, aus Norwegen und aus Schweden an die entsprechenden nationalen Resolver sowie Anfragen zu persistenten Identifikatoren aus den DOI-Schemata (Digital Object Identifier), Handle und Ark (Archival Resource Key) an die jeweiligen Dienste weiter. Der Resolver wurde im Rahmen von EuropeanaConnect entwickelt und ist auch Teil der Infrastruktur der Europeana.

Der Resolver der Deutschen Nationalbibliothek verwaltet nicht nur persistente Identifikatoren für digitale Pflichtexemplare, sondern jeder Anbieter digitaler Inhalte kann persistente Verknüpfungen zu den Objekten hinterlegen. Im Mai 2011 verwendeten über 400 Institutionen diese Möglichkeit und hatten knapp 5 Millionen URNs registriert. Mit etwa 3.500 Anfragen pro Tag sind die Zugriffsraten noch recht moderat, es ist aber zu erwarten, dass diese Zahlen mit zunehmender Verwendung von persistenten Identifikatoren steigen werden.

Die Deutsche Nationalbibliothek archiviert alle abgelieferten digitalen Objekte in einem Repository. Dieses Repository für digitale Objekte ist zuständig für die Bereitstellung von Arbeitskopien der Publikationen; dies kann eine Kopie der Original-Publikation sein oder – insbesondere im Fall digitalisierter Exemplare – ein Derivat wie z. B. eine JPEG-Kopie eines TIFF-Bildes. Wenn eine Netzpublikation im Repository abgeliefert wird, erstellt ein automatisierter Prozess ein Datenpaket für die digitale Langzeitarchivierung und übermittelt es an ein separates Archivsystem, das für die Langzeitarchivierung des Objekts zuständig ist.

Die Kombination von digitaler Langzeitarchivierung und persistenten Identifikatoren bietet viele Vorteile. Wissenschaftler können mit Hilfe der persistenten Identifikatoren in ihren Arbeiten Referenzen auf digitale Objekte in der Gewissheit geben, dass die Referenz auch über lange Zeit stabil bleibt und somit auch andere Wissenschaftler ohne weiteren Aufwand die zitierte Ressource abrufen können, um Aussagen zu verifizieren. Verlage und andere Dokumentenhersteller können von der Deutschen Nationalbibliothek schon vor der Veröffentlichung einer Publikation einen persistenten Identifikator beantragen und diesen vorab im Dokument aufführen, damit dieser z. B. auch in Druckfassungen des Dokumentes verfügbar ist. Der Einsatz des Resolvers ermöglicht es Verlagen oder Datenverwaltern auch,

äquivalente Objekte an verschiedenen Orten zu speichern und dem Resolver dann anzugeben, welche Adresse die höchste Priorität hat: z. B. kann ein Verlag dieselbe URN nutzen, um eine Publikation zu identifizieren, die sowohl im Online-Shop des Verlags als auch im Repository der Bibliothek verfügbar ist. Bei der Auflösung der URN wird die Adresse des Objekts im Online-Shop die höhere Priorität haben, so dass der Nutzer zum (kommerziellen) Angebot des Verlags geleitet wird. Sollte der Verlag aus irgendeinem Grund aus dem Geschäft aussteigen oder den Link einfach nicht weiterpflegen, wird die Referenz zum digitalen Archiv der Bibliothek immer noch verfügbar sein.

Digitale Langzeitarchivierung ist eine verantwortungsvolle Aufgabe, und es ist wichtig, sicherzustellen, dass die Daten so weit wie möglich gegen Veränderungen und Einwirkungen von außen geschützt werden. Das digitale Archiv, das derzeit an der Deutschen Nationalbibliothek implementiert ist, basiert auf der Software DIAS, einem System, das im Projekt KOPAL in Zusammenarbeit zwischen der Deutschen Nationalbibliothek, der Niedersächsische Staats- und Universitätsbibliothek Göttingen und IBM Deutschland entwickelt wurde. Schon in diesem Stadium ist offensichtlich, dass es notwendig sein wird, klassische Langzeitarchive wie DIAS durch eine international vernetzte Archivierungsinfrastruktur einer neuen Generation zu ersetzen, in der Dokumente redundant in mehreren, untereinander verbundenen Langzeitarchiven abgelegt werden. Die Deutsche Nationalbibliothek ist am EU-Projekt SHAMAN (Sustaining Heritage Access through Multivalent Archiving) beteiligt, in dem technische und konzeptionelle Grundlagen für eine neue Generation vernetzter Systeme entwickelt werden sollen. Die Langzeitarchive sollen mit Hilfe von GRID-Technologien verbunden werden, um die Bewältigung komplexer und ressourcenintensiver Aufgaben der Langzeitarchivierung zu ermöglichen. Um künftigen Bedürfnissen nach einer flexiblen und skalierbaren IT-Infrastruktur gerecht zu werden, rüstet die DNB derzeit ihr Rechenzentrum auf, um sicherzustellen, dass wir den Speicherplatz für das digitale Repository und das Langzeitarchiv leicht erweitern können sowie weitere Serverknotenpunkte einbauen können, um ausreichend Rechenleistung zur Verfügung zu stellen.

Langzeitverfügbarkeit bedeutet nicht nur, die Dokumente unverändert zu erhalten, sondern setzt auch voraus, dass sie auch in Zukunft nutz- und lesbar bleiben. Während derzeitige, digitale Archive, wie z. B. DIAS, nur die Bitfolge aufbewahren, ist sich die Deutsche Nationalbibliothek dessen bewusst, dass die Herausforderung darin besteht, sicherzustellen, dass das Objekt auch mit künftigen Betriebssystemen und den Anforderungen künftiger Softwaregenerationen nutzbar ist.

Im Rahmen des EU-Projekts KEEP (Keeping Emulation Environments Portable) untersuchen die Deutsche Nationalbibliothek und andere Partner, wie eine möglichst genaue Wiedergabe von statischen sowie dynamischen Objekten erreicht werden kann. Dies umfasst alle Arten von Objekten: Text, Sound, Bilder, Multimediadokumente, Webseiten, Datenbanken, Videospiele u. a. Ziel des Projektes ist es, flexible Zugriffswerkzeuge und Speichermöglichkeiten zu entwickeln und so die Langzeitverfügbarkeit zu gewährleisten. Wenn diese Tools etabliert und die Erhaltungsmaßnahmen an den Objekten in den digitalen

Langzeitarchiven durchgeführt worden sind, werden separate Prozesse sicherstellen, dass das migrierte Objekt zum digitalen Repository zurückgespielt wird, um dem Benutzer Zugriff auf die aktuellste Version zu gewährleisten.

Herausforderungen, Probleme und Chancen

Zu den Herausforderungen bei der Sammlung von Netzpublikationen gehören natürlich die sich überaus schnell verändernden technischen Gegebenheiten – nicht nur die Publikationsform verändert sich ständig durch den Einsatz neuer Technologien, sondern auch die Möglichkeiten der Übermittlung elektronischer Daten und ihr Einsatz werden laufend weiterentwickelt. Es ist gar nicht so lange her, dass die Möglichkeit, Texte elektronisch zu speichern, zu verändern und am Computer zu bearbeiten, als Fortschritt galt. Danach folgten rasch die Variabilität der Darstellung, die Einbindung multimedialer Inhalte und die Möglichkeit, Werke zu erarbeiten, bei denen letztlich keine Endfassung erkennbar ist, die mit den traditionellen Erscheinungsformen vergleichbar wäre. Da der Auftrag der DNB nicht nur die Sammlung, sondern auch die Bereitstellung und vor allem die Archivierung der Sammlung umfasst, benötigen wir weitreichende Lösungen, um die durch die Formate gesetzten Grenzen zu überwinden.

Praktische Probleme gibt es bereits bei derzeit gut etablierten Verfahren, bei denen Texte in einem internen System gespeichert sind und erst durch die Aktion des Benutzers zu einem Werk zusammengefügt werden. Jedem sind Nachschlagewerke wie Wikipedia, aber auch Fachdatenbanken bekannt, deren Inhalte nicht ohne die im Hintergrund laufende Software sinnvoll genutzt werden können.

Auch neue Leseformate stellen die Bibliothek vor neue Herausforderungen – oft sind die Formate, in denen die Dateien gespeichert werden, nur mit entsprechenden Lesegeräten zu öffnen und mit Verschlüsselungseigenschaften versehen, die weder eine allgemeine Zugänglichkeit noch eine Archivierung erlauben.

Neben den Herausforderungen, die sich aus den neuen Techniken ergeben, müssen wir uns mit Metadatenaspekten befassen. Im Laufe der letzten Jahre wurden viele Gespräche mit Verlegern geführt. Dabei hat sich herausgestellt, dass für Bibliotheken – da sie an den Austausch von Metadaten und die Anwendung von Standards in diesem Prozess gewohnt sind – die gemeinsame Nutzung von Metadaten selbstverständlich geworden ist. Verlage und alle Unternehmen, die Produkte verkaufen, haben eine andere Sichtweise auf Metadaten: für sie sind die Daten hauptsächlich ein Mittel ihr Geschäft voranzubringen, während Bibliotheken andere Anforderungen haben – insbesondere diejenigen, die den Auftrag haben, die Publikationen eines Landes zu verzeichnen. Diese verschiedenen Sichtweisen auf Metadaten können die Harmonisierung erschweren. Ein reibungsloser Datenaustausch ist nur zwischen Beteiligten möglich, die dieselben Erwartungen haben und die allgemein anerkannte Standards nutzen, die auf demselben Verständnis von Metadaten beruhen, z. B. zwischen Bibliotheken oder auch mit nationalen Buchhandelsorganisationen, die das Erstellen von Metadaten als hochwertige Tätigkeit ansehen.

Nächste Schritte

Zusätzlich zu den monografischen Netzpublikationen werden elektronische Zeitschriften und Zeitungen gesammelt. Seit Mai 2010 werden in einem DNB-Projekt mit Unterstützung eines Dienstleisters 300 Tageszeitungen laufend eingesammelt, im Katalog verzeichnet und im Repository archiviert. Die Deutsche Nationalbibliothek konzentriert sich nun auf die Sammlung von E-Journals. Eine Besonderheit der E-Journals gegenüber den E-Books ist, dass sie individuelle Metadatenformate nutzen, da es für elektronische Zeitschriften keine allgemein verwendeten Metadatenstandards gibt. Eine Übertragung von verfügbaren Metadaten wird dadurch erschwert oder nahezu unmöglich gemacht. Deshalb müssen alternative Wege für ihre Erschließung gefunden werden.

In den letzten Jahren wurden von deutschen Bibliotheken in großem Umfang gedruckte Bücher und Zeitschriften digitalisiert. Soweit sie öffentlich zugänglich sind, fallen diese unter die Ablieferungspflicht. Zur Verbesserung der Archivierung strebt die Deutsche Nationalbibliothek an, nicht die für die Nutzer bereitgestellten Datenformate zu sammeln, sondern vielmehr, wenn möglich die Masterdateien zu erhalten, die eine verlustfreie Speicherung ermöglichen. Die Deutsche Nationalbibliothek führt Gespräche mit den Bibliotheken und öffentlichen Einrichtungen, die Erfahrungen mit der Digitalisierung haben, um diese Aufgabe kooperativ zu bewältigen.

Websites werden derzeit noch nicht gesammelt. Allerdings arbeitet die Deutsche Nationalbibliothek im Rahmen eines Projekts daran, auch einen entsprechenden Workflow für die Sammlung, Archivierung und Erschließung von Websites einzurichten.

Das Spektrum der Dateiformate für die digitalen Objekte muss erweitert werden, damit der Ausschnitt an Netzpublikationen, den wir langfristig erhalten wollen, erweitert werden kann. Dafür muss der Umgang mit Material, das mit Verschlüsselungsmechanismen geschützt ist, in Bezug auf seine Langzeitarchivierung und –verfügbarkeit getestet werden.

Unser Ziel ist es, weitere Möglichkeiten für die Ablieferung von Netzpublikationen zu schaffen, um es mehr Verlagen zu ermöglichen, über mindestens eine der Schnittstellen abliefern zu können. Die entsprechenden Mappings werden vorbereitet und müssen zwangsläufig auch ständig aktualisiert werden.