

Zur Archivierung Digitaler Amtlicher Druckschriften – Überlegungen im Kontext der Bibliotheken

Reinhard Altenhöner

Amtliche Druckschriften oder Amtsdrucksachen liegen im Schnittpunkt der Aktivitäten von Archiven und Bibliotheken. Sie sind so ein geeignetes Beispiel für die Beschreibung der gemeinsamen Probleme, denen sich Organisationen, die auf den Erhalt des kulturellen Erbes verpflichtet sind, stellen müssen. Dies soll im vorliegenden Beitrag aus der Perspektive der Bibliotheken und ihrer konkreten Aktivitäten im weiteren Zusammenhang der elektronischen Publikationen geschehen, um daran einige Überlegungen zu gemeinsamen Handlungsfeldern von Archiven und Bibliotheken anzuschließen. Dabei stehen Anforderungen im Umfeld der Langzeitarchivierung im Vordergrund.

1. Amtliche Druckschriften

Amtliche Druckschriften sind Veröffentlichungen staatlicher oder halbstaatlicher Stellen, die für die öffentliche Verbreitung bestimmt sind. Als Urheber fungieren in aller Regel Behörden und Körperschaften des öffentlichen Rechts, also der Staat mit seinen verschiedenen Verwaltungsorganen bis hinab zu den Gemeinden und Volksvertretungen. Beispiele für Amtliche Druckschriften sind demnach Gesetzblätter, Haushaltspläne, statistische Jahrbücher, aber auch ein-

zelne Gesetze, Verordnungen oder Entscheidungen.¹ Hinzu kommen aber auch Informationsschriften aller Art, die zwar nicht im Buchhandel erscheinen, sich aber zumeist an eine breite Öffentlichkeit richten, aufklärerischen oder informatorischen Zwecken dienen und häufig den Charakter eines Rechenschaftsberichtes der jeweiligen herausgebenden Institution haben. Damit sind sie aber als Quellen zumindest von rechts- und staatswissenschaftlichem Interesse – insbesondere auch in historischer Perspektive.

In gewissem Sinne stellen Amtliche Druckschriften so das Gegenstück eines Verwaltungsvorgangs dar, der ja im Allgemeinen nur den unmittelbar Beteiligten bekannt ist – und der in der Regel den Kern der archivalischen Überlieferung bildet. Andererseits sind Amtliche Druckschriften aber wiederum häufig selbst das Ergebnis eines solchen oft länger andauernden Prozesses. Sie bilden – obschon traditionell auch Gegenstand bibliothekarischer Sammlungsanstrengungen – als Objekt in archivalischen Sammlungen einen wichtigen Arbeitsbereich von Archiven. Dabei gelangen sie in publizierter Form als Teil oder Ergebnis eines Vorgangs auch in die archivierungsrelevante Überlieferung des Verwaltungshandelns und werden auf diesem Weg Teil des Archivgutes. Typisch dafür und für die Systematik der Erschließung, über die im Archivbereich solche Informationen bereitgestellt werden, ist die vorgangsbezogene Ablage bzw. Zusammenführung von Quellenmaterialien nach Vorgang.

Damit sind die Amtlichen Druckschriften allerdings in den Archiven oftmals schwer als Einzelobjekt identifizierbar. Für die Bibliotheken hingegen steht die Publikation im Mittelpunkt, das Umfeld, aus dem heraus etwa eine Publikation entsteht, ist im Allgemeinen irrelevant, es sei denn, es hätte selbst wiederum zu Publikationen geführt. Eine relativ große Zahl der Amtlichen Druckschriften erscheint regelmäßig, die einzelnen Objekte stehen also miteinander in einer logischen Verknüpfungsbeziehung, die auch über die Erschließung erkennbar und nachvollziehbar sein sollte.

Auch wenn die Sammlung und Erschließung der Amtlichen Druckschriften grundsätzlich geregelt ist, werden sie in Deutschland im Kontext der Bibliotheken eher stiefmütterlich behandelt: Zwar bilden sie eine eigene Sammlungsgruppe und werden häufig auch als solche gekennzeichnet; sie werden aber in vielen Bibliotheken, die das Pflichtexemplarrecht wahrnehmen, nur in Auswahl bzw. ohne Anspruch auf Vollständigkeit gesammelt. Aufgrund eines jüngst neu gefassten Erlasses ist auf Bundesebene festgelegt worden, dass neben der Deutschen Nationalbibliothek eine Reihe von weiteren national bedeutsamen Bibliotheken (Bundestagsbibliothek, Staatsbibliothek zu Berlin und die Bayerische Staatsbibliothek) das entsprechende Pflichtexemplar erhalten. Ergänzt wird diese nationale „Verteilungsebene“ durch weitere Regelungen auf Landesebene, in denen die Landes- und Regionalbibliotheken zur Sammlung verpflichtet werden; hinzu tritt dann noch – in der Übersicht schwer greifbar – die kommunale Ebene. Konkret bedeutet dies für die erstellende Einrichtung, dass sie bis zu fünf und unter Umständen mehr Exemplare ihrer Publikation an die jeweils Berechtigten abliefern muss. Die Intensität, in der dieser Auftrag von der erstellenden Einrichtung wahrgenommen wird, schwankt; Gleiches gilt für die Auslegung der Regeln im Detail. So sammelt die Deutsche Nationalbibliothek – in weiter Auslegung ihres Auftrags – auch Publikationen der Kirchen oder anderer „halbamtlicher“ Stellen. Hinzu kommen NfD- und Verschlussachen mit teilweise oder ausschließlich amtlichem Inhalt, ferner Werke von Kreisen und Kommunen mit ausschließlich amtlichem Inhalt.

¹ Vgl. Rehm, Margarete: Lexikon Buch Bibliothek Neue Medien. München 1991, S. 8f. (Artikel amtliche Druckschriften, Amtsdruksachen).

Auf der Ebene der Länder ist die Situation insgesamt noch offener: der Auftrag, die entsprechenden Schriften zu sammeln, wird im Allgemeinen nicht vollständig und umfassend wahrgenommen, sondern sehr selektiv und mit reduziertem Aufwand. Es sei noch ergänzt, dass die Staatsbibliothek zu Berlin den Bereich der Internationalen Amtlichen Druckschriften in ihrem Sammelauftrag verankert hat, also Veröffentlichungen staatlicher und öffentlich-rechtlicher Institutionen sowie internationaler – zwischenstaatlicher – Organisationen.²

Die Archive verfügen insgesamt über umfangreiche Sammlungen von Amtlichen Druckschriften, so beispielsweise das Bundesarchiv aus dem Bereich des Militärs.³

Welche Dimension die Amtlichen Druckschriften rein mengenmäßig haben, ist kaum auszumessen und zumindest in den auf nationaler Ebene hier aktiven Einrichtungen nicht genau bekannt. Für den Erscheinungszeitraum 1947-2001 verzeichnen die Nachweissysteme der Deutschen Nationalbibliothek rund 170.000 Druckschriften, allerdings ist eine Hochrechnung dieser Zahlen auf den Gesamtbestand aufgrund der unvollständigen und derzeit nicht mehr durchgeführten spezifischen Kennzeichnung der Bestände nicht stabil.

2. Elektronisch publizierte Amtliche Druckschriften

Ganz unzweifelhaft wächst die Zahl der auch oder nur noch online über das Internet bereitgestellten Publikationen deutlich an. Und gleichzeitig erweitert sich der Bereich dessen, was als Publikation bezeichnet wird, ebenfalls deutlich: Wenn etwa früher eine Einrichtung in einer Imagebroschüre über ihre Aufgaben und Funktionen informierte, so tut sie dieses heute in einer viel größeren Tiefe und mit gleichzeitig sehr viel höherem Aktualitätsanspruch auf ihrer Website. Diese Änderung ist ganz sicher auch quantitativ nachvollziehbar. Der Aufwand für die Publikation ist – zumindest hinsichtlich einer Relation Informationseinheit zu Ressourceninvestition – deutlich gesunken und damit auch die Schwelle, über die gegangen werden muss, wenn man publizieren will. Diese Beobachtungen gelten in gleicher Weise auch für die Amtlichen „Druckschriften“; besser sprechen wir vielleicht nur noch von Amtlichen Publikationen generell bzw. elektronischen Amtlichen Publikationen oder Amtlichen Netzpublikationen. Aufgrund der Nomenklatur im neuen Gesetz zur Deutschen Nationalbibliothek (DNBG) benutze ich im Folgenden den Begriff der Amtlichen Netzpublikationen.⁴

Ganz sicher jedenfalls kann die elektronische Publikationsform zunehmend zumindest potentiell auch viele konventionelle Druckschriften – und hier insbesondere die regelmäßig sich wiederholenden Berichte – ersetzen. Zwar sind noch für viele Bereiche Printausgaben erforderlich (obschon parallel eine Onlinversion zur Verfügung gestellt wird, beispielsweise die Publikation von Bundesgesetzen), um zum Beispiel ein Gesetz zu verkünden und damit in Kraft zu setzen, aber auch hier dürfte mittelfristig eine Veränderung anstehen und die Zahl der „born-digital-Publikationen“, die gar nicht mehr in gedruckter Form erscheinen, dürfte steigen.

² Vgl. <http://amtsdruckschriften.staatsbibliothek-berlin.de/de/links/internetquellen/hinw.html>. (erreichbar im Juni 2006). Zum Schriftentausch vgl. Metz, Johannes: Internationaler Amtlicher Schriftentausch (65th IFLA Council and General Conference Bangkok, Thailand, August 20 - August 28, 1999) vgl. <http://www.ifla.org/IV/ifla65/papers/054-83g.htm> (erreichbar im Juni 2006).

³ http://www.bundesarchiv.de/aufgaben_organisation/abteilungen/ma/00878/index.html.

⁴ Vgl. den Entwurf der Bundesregierung unter http://www.bundestag.de/ausschuesse/a22/berichte/nationalbibliothek/drs__16_322.pdf und http://www.bundestag.de/ausschuesse/a22/berichte/nationalbibliothek/deutsche_nationalbibliothek.pdf (erreichbar im Juni 2006).

Für Bibliotheken wie für Archive bedeutet diese Entwicklung einen gravierenden Wandel: Wenn früher ein formal definiertes bzw. definierbares Druckobjekt im Mittelpunkt stand, so war dieses in seinen Eigenschaften und (Markt)Potentialen klar beschrieben. Heute aber handelt es sich häufig um ganze Informationscluster, die auf Einzeldateiebene nicht mehr nutzbar bzw. greifbar sind. Nicht nur für ihre Übernahme in die Sammlungen, sondern auch bei ihrer Erschließung treten erhebliche Vermittlungs- und Akzeptanzprobleme auf, die sich letztlich wiederum massiv auf die Zugriffsmöglichkeiten auswirken.

Betrachtet man nun konkret den Anteil digitaler Publikationen in der Gesamtzahl der Amtlichen Publikationen, wird deutlich, dass die Zahl elektronischer Publikationen als Surrogat einer Printausgabe im deutschsprachigen Bereich langsam erkennbar zunimmt, sehr häufig als Parallelpublikation (oft auch rückwirkend!) zu einer gedruckten Version. Allerdings sind (auch hier) keine genauen Zahlen bekannt, die Bayerische Staatsbibliothek stellt nur fest, dass es viele seien;⁵ in der Deutschen Nationalbibliothek sind ebenfalls keine exakten Zahlen bekannt. Die Zahl originär elektronisch publizierter Dokumente („born-digital“) ist nach der (subjektiven) Einschätzung der zuständigen Fachabteilungen der Deutschen Nationalbibliothek allerdings bislang gering.

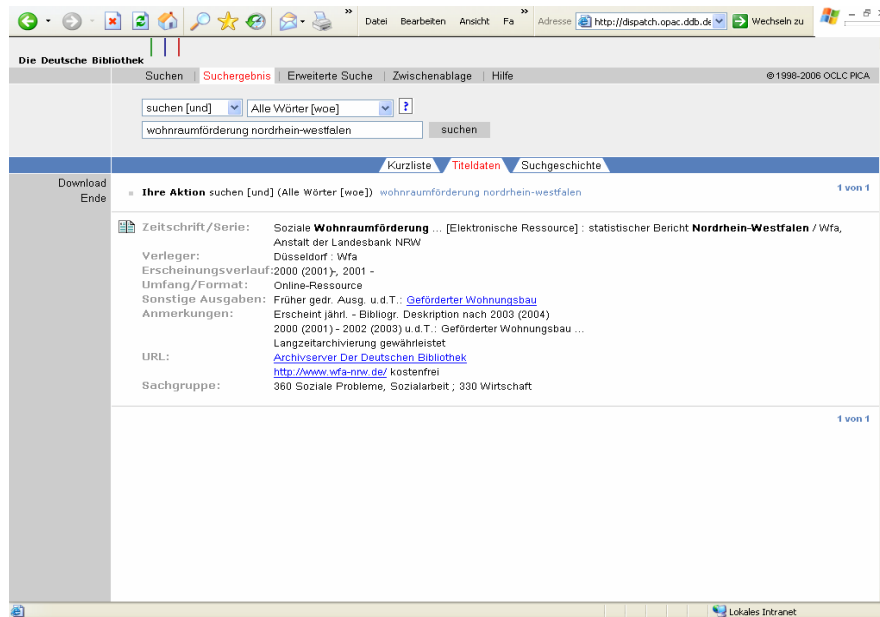
Das große Potential und die Dynamik, die im E-Publishing liegt, wird jedoch dazu führen, dass die Zahl dieser Publikationen deutlich ansteigen wird und genau vor diesem Hintergrund stellt sich die Frage nach der dauerhaften Vorhaltung dieser Publikationsgattung in elektronischer Form um so intensiver. Aus heutiger Sicht ist für Deutschland festzuhalten, dass bislang die systematische Sammlung elektronischer Amtspublikationen nicht klar und abschließend verbindlich organisiert ist. Das hängt zum Teil auch damit zusammen, dass Abgrenzungs- und Eingrenzungsprobleme und daraus folgend auch technisch-operative Probleme bestehen: So werden die Publikationen zum Beispiel über das Web-Angebot einer Institution eingebunden oder auch integriert angeboten mit der Folge, dass vielleicht selbst bei einer weiten Begriffsdefinition der komplette Webauftritt einer Einrichtung eine Amtliche Publikation darstellt. Aus dieser Situation heraus arbeitet die Bundestagsbibliothek daran, ein Archiv der Web-Seiten der Bundesregierung aufzubauen;⁶ auch die Deutsche Nationalbibliothek hat zu diesem Thema mit einer Vorher-/Nachhersammlung rund um die letzte Bundestagswahl 2005 in Zusammenarbeit mit dem European Internet Archive experimentiert.⁷ Angesichts der großen Menge der zu sammelnden Daten und der noch nicht stabilen technischen Lösungen ist die Validität der gesammelten Daten bislang allerdings weder quantitativ noch qualitativ bezogen auf das Gesamtspektrum der Aufgabe überzeugend.

Eine weitere Aufgabe besteht darin, die Sammlung von „Zeitschriften“ oder „Serien“-ähnlichen Publikationen, also fortlaufend unter einem Titel in Einzelstücken erscheinenden Publikationen, zu organisieren. Diese führen gewissermaßen permanent zu „neuen Fällen“, die bearbeitet werden müssen. Die dafür geeigneten automatischen Abläufe sind komplex und müssen zum Teil erst entwickelt werden. Es gibt keine zentrale Instanz, die staatliche Publikationen (wenigstens auf Bundesebene) systematisch veröffentlicht und auch die Ablieferung und Archivierung organisieren würde. Auch die persistente Adressierung der Objekte im Internet stellt eine Herausforderung dar.

⁵ http://www.bsb-muenchen.de/Amtsdruckschriften_der_BRD.188.0.html?&L=&styl= (erreichbar im Juni 2006).

⁶ http://www.langzeitarchivierung.de/downloads/dresden2006_ullmann.pdf (erreichbar im Juni 2006) und den Beitrag „Heute im Netz – morgen im Archiv“ auf der 10. Tagung des Arbeitskreises „Archivierung von Unterlagen aus digitalen Ebenen“ in diesem Band.

⁷ <http://europarchive.org/> (erreichbar im Juni 2006).



Beispiel einer elektronischen Ressource (Amtliche Netzpublikation) im WEB-Katalog der Deutschen Nationalbibliothek

Obwohl die anstehenden Aufgaben recht klar sind, ist es bislang in Deutschland nicht gelungen, ein abgestimmtes Vorgehen für den Gesamtkomplex der Amtlichen Publikationen zu erreichen. Neben den genannten Problemen „leiden“ die amtlichen Publikationen auch an dem Stigma, von mangelnder Relevanz zu sein. Dieses Bild wendet sich aber vielleicht gerade auf der nationalen Ebene, denn international werden Amtliche Publikationen in letzter Zeit stärker beachtet, was auch ein größeres, von der Library of Congress (LoC) initiiertes Projekt belegt.

3. Internationale Bemühungen zur Optimierung der Sammlungsaktivitäten / Das Projekt IEX (International Exchange)

Ziel des Projekts ist es, die internationale Zusammenarbeit für diese Publikationsgattung deutlich zu intensivieren und durch den automatisierten Austausch von Metadaten und Objekten insbesondere den internationalen Austausch zu vereinfachen. Für ihre nationale Sammlung hat die Library of Congress (LoC) einen Arbeitsablauf mit dem Government Printing Office organisiert und sie sucht – beginnend mit Deutschland – nach internationalen Partnern für die Aufnahme von (automatisierten) Tauschbeziehungen. Einbezogene Einrichtungen sind neben der Deutschen Nationalbibliothek die Bayerische Staatsbibliothek und die Staatsbibliothek zu Berlin. Für den Bereich der elektronischen Serien wurde – auch aufgrund der Zusammenarbeit im Kontext der Nutzung der Elektronischen Zeitschriftenbibliothek für Lizenzinformationen – die Universitätsbibliothek Regensburg miteinbezogen, für bestimmte technische Fragen auch das Medien- und Servicezentrum der Berliner Humboldt-Universität.

Für den Bereich der in der Zeitschriftendatenbank (ZDB) bzw. der Elektronischen Zeitschriftendatenbank (EZB) verzeichneten Zeitschriften und Serien konnte die Zusammenarbeit bereits konkret gestartet werden, indem ZDB und EZB der LoC gemeinsam entsprechende Metadaten zur Verfügung stellen. Bei allen anderen Bereichen erweist sich die problematische Organisationslage in

Deutschland für den Bereich der Amtlichen Netzpublikationen als ausgesprochen hinderlich.

Weitere Schritte werden derzeit vor allem im Aufbau einer technischen Struktur gesehen; zur Diskussion steht so zum Beispiel, Ergebnisse des Projekts LOCKSS (Lots of Copies Keep Stuff Safe)⁸ zu nutzen und als Austausch- und Bereitstellungsplattform für Metadaten und Objekte einzusetzen.

Ein weiteres gravierendes Problem wird unter den Projektbeteiligten in der Langzeitverfügbarkeit der archivierten Bestände gesehen. Die Langzeitarchivierung und die dauerhafte Bereitstellung der Objekte in einer nachnutzbaren Form stellen tatsächlich eine der großen Herausforderungen der aktuellen Informationsgesellschaft dar – auch für LOCKSS wird dieses Thema zunehmend bedeutsamer, weil das Teilnehmerkonsortium inzwischen hier Anforderungen stellt, die über das mehrfache verteilte Spiegeln von Daten auf einfacher und daher günstiger Infrastruktur hinausgehen und eine längerfristige Perspektive für die Bereithaltung der Daten und ihrer Nutzung einfordern.

Demgegenüber ist das deutsche Projekt kopal (Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen)⁹ schon vergleichsweise weit gediehen, es soll im Folgenden – nach einer Einführung in die Problematik – ausführlich vorgestellt werden.

4. Herausforderungen für ein Archiv: Langzeitverfügbarkeit

Heute entstehende Publikationen liegen in aller Regel originär elektronisch vor; die Produktions- oder Erstellungssysteme basieren auf der Informationstechnik. Ob die Publikation dann als „born-digital“-Veröffentlichung im WWW zugänglich gemacht wird oder ob für die primäre Verbreitung eine Druckausgabe erstellt und in den Handel gegeben wird, ist letzten Endes nicht entscheidend – immer bilden digitale Informationen die ursprüngliche Quelle. Und in noch umfassenderem Sinne gilt dies für (Roh)Daten, die aus der experimentellen oder beobachtenden Forschung stammen, also zum Beispiel Daten von Satelliten, aus Großforschungsanlagen, aber auch beispielsweise aus Geoinformationssystemen. Für alle diese Datengruppen gilt, dass hier bereits heute erhebliche Aufwände erbracht werden müssen, um dieses Material zu sichern und kurz- und mittelfristig für Wissenschaft und Forschung zur Verfügung zu halten.

Die Problematik wird in der Rückwärtsbetrachtung deutlich: So bestand etwa bei einer größeren Gruppe von Datenträgern (Disketten) aus der ersten Heimcomputergeneration, die sich heute in der Sammlung der Deutschen Nationalbibliothek befinden, das ernsthafte Risiko, aufgrund von Alterungsschäden an den Datenträgern die archivierten Informationen zu verlieren. Konkret bedeutet dies, dass die Datenträger schlicht nicht mehr genutzt werden konnten, weil sie selbst Fehler aufwiesen, zunehmend aber auch, weil die Lesegeräte – in diesem Fall die Homecomputer der ersten Generation aus den achtziger Jahren – nicht mehr brauchbar sind. Doch selbst nachdem nach Ankauf alter Hardware diese Situation durch den Transfer der Daten als ISO-Image auf einen anderen Träger erfolgreich entschärft werden konnte, erwies es sich als Problem, dass die nun gesicherten Daten nicht ohne weiteres in einer MS-Windows basierten Anwendungsumgebung ablauffähig waren. Hier mussten letztlich spezielle Softwaretools eingesetzt werden, die es erlaubten, die alte Umgebung so zu emulieren,

⁸ <http://www.lockss.org/lockss/Home> (erreichbar im Juni 2006).

⁹ „Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen“; Projekt-Website <http://kopal.langzeitarchivierung.de/> (erreichbar im Juni 2006).

dass auch in einer aktuellen Anzeigeumgebung die alten Formate nutzbar wurden. Dieses konkrete Projekt, das eine im Vergleich zur Gesamtmenge der in der Deutschen Nationalbibliothek auf physischen Trägern vorgehaltenen Materialien geringen Teilbestand von wenigen Tausend Objekten betraf, erhellt die Ausgangssituation exemplarisch. Die Erkenntnisse hieraus lassen sich generalisieren:

Obschon die Menge der bislang durch das Obsoletwerden von Formaten kritisch gewordenen Daten relativ gering und insgesamt die Zahl der Verluste klein ist (überwiegend eingetreten im Bereich der Primärdaten), ist die kurze Geschichte der Informationstechnik von so vielen Formatwechseln und darauf aufsetzenden Editor-Generationen geprägt, dass allein schon die Verwaltung dieser Menge an möglichen Hardware-/Softwarekonfigurationen kritisch ist. Nimmt man hinzu, dass in aller Regel kein Editor nativ auf einer Hardwarestruktur aufsetzt, sondern wiederum Software voraussetzt – häufig in spezifischen Versionen – die wiederum auf einer speziellen Hardwarekonfiguration basiert, wird deutlich, wie viele mögliche Konstellationen abgefangen werden müssen. Dreh- und Angelpunkt dafür sind daher Metadaten, die die historische Umgebung, aus der heraus ein Publikationsobjekt entstanden ist, nachvollziehbar und vor allem (maschinell) nachnutzbar beschreiben.



Abbildung der verschiedenen Lese-/Trägerebenen bei der Nutzung einer elektronischen Ressource

Folgende grundsätzliche Ausgangsbedingungen und auch international akzeptierte Grundsätze sind erkennbar:

- Eine Grundbedingung für alle Bemühungen zur langfristigen Sicherung des Zugangs auf digitale Information ist der Erhalt des binären Datenstroms des datentechnischen Ausgangsmaterials für die Übermittlung digitaler Informationen. Die dazu erforderliche Technologie ist im Prinzip bekannt und vorhanden und wird seit Jahrzehnten von Rechenzentren erfolgreich eingesetzt. Methodisch gehören zu diesen Datensicherungsmaßnahmen insbesondere regelmäßige Überprüfungen des gespeicherten Materials und die systematische Überführung auf neue Datenträger, wenn dies durch Technologiewechsel angezeigt ist.

- Davon ausgehend, dass der originär gespeicherte Bitstream zur Verfügung steht und dies im Übrigen auch nachweisbar überprüft werden kann (Authentizität der Daten!), benötigt der Anwender in der Regel eine spezifische Software- (und ggf. Hardware)umgebung, um das digitale Objekt, das ja in einem bestimmten Format gespeichert wurde, anzuzeigen und zu nutzen. Durch den schnellen Technologiewechsel allerdings ist der Zugriff auf ältere Datenformate zunehmend erschwert und das Wissen um die Eigenschaften der historischen Formate geht verloren.

In der internationalen Diskussionen zeichnen sich im Wesentlichen zwei Strategien ab, die als wichtige Lösungsstrategien akzeptiert werden: die Migration und die Emulation.

- Gegenstand der **Migration** ist die rechtzeitige Konversion alter Dateiformate in ein aktuell nutzbares Format. Dies muss natürlich zu einem Zeitpunkt erfolgen, zu dem die Formate noch vollständig und verlustfrei gelesen werden können. Eine solche Überführung aus bestimmten Hardware- und Software-Konfigurationen in eine neuere, technisch und organisatorisch beherrschte Umgebung geschieht im Idealfall periodisch wiederkehrend, wobei es sich empfiehlt, Ausgangsversionen zumindest zeitweise zu erhalten, um so einen „Rückkehrpunkt“ zu haben, wenn sich herausstellt, dass die konvertierte Datei fehlerhaft ist. Zweck der Migration ist es also, die Integrität und die Verfügbarkeit digitaler Ressourcen trotz des stetigen Wandels der technischen Umgebung zu erhalten. Problematisch bei diesem Verfahren ist der möglicherweise nicht erkannte Verlust von (Teil-)Informationen bei automatisch ablaufenden Migrationsroutinen; je komplexer das vorliegende Datenformat ist, desto größer wird die Möglichkeit, dass bei längeren Migrationsketten (über Jahre hinweg) bestimmte Features nicht mehr vorhanden bzw. lauffähig sind. Auf der anderen Seite aber ist die Migration ein geeignetes Verfahren für große Datenmengen, und sie dürfte daher vor allem für statische Formate die vernünftigste Strategie sein.
- Bei der **Emulation** hingegen wird mit einem oder ggf. mehreren speziellen Programmen (dem Emulator) eine alte Systemumgebung, die zur Zeit der Entstehung des jeweiligen Formates üblich war, in einer aktuellen Systemumgebung nachgebildet. Auf einem solchen System werden also Daten und Programme verarbeitet, die ursprünglich für ein anderes, historisches System bestimmt waren. Dies kann unter Umständen sehr aufwändig sein – ggf. sind nämlich mehrere ineinander verschachtelte Emulationsebenen oder -instanzen erforderlich – und setzt vor allem eine sehr genaue Definition der hard- und softwareseitigen Systemanforderungen voraus. Allerdings erlaubt es diese Methode, auch sehr komplexe Formate (z.B. Multimedia-Anwendungen) dauerhaft zu nutzen, die sich aufgrund der mächtigen und oft proprietären Funktionen, die über das jeweilige Format und die spezifische Software angeboten werden, einem Migrationsverfahren entziehen.

Beide Herangehensweisen, die Migration und die Emulation, sollten aufgrund ihrer je spezifischen Möglichkeiten je nach konkreter Aufgabenstellung im Kontext eines Archivsystems zur Langzeiterhaltung und Langzeit-Verfügbarmachung digitaler Publikationen einsetzbar sein. In einem Archivsystem muss es also möglich sein, sowohl Format-Migrationen im Sinne regelmäßiger Konversionsläufe durchzuführen, als auch die notwendigen Informationen für den Aufbau einer je geeigneten Emulationsumgebung vorzuhalten und bereitzustellen. Nach dem Gesagten dürfte klar sein, dass für den Einsatz beider Strategien Metainformationen zum Format und den erforderlichen Hard- und Softwarebedingungen bis hin zum technischen Entstehungskontext eines Objekts erforderlich sind. Nur aufgrund solcher Informationen lassen sich erfolgreich Strategien

zur Sicherung der Langzeitverfügbarkeit auf digitale Objekte wie zum Beispiel Amtliche Netzpublikationen aufsetzen. In einem Archivsystem, das unter diesen Ausgangsbedingungen arbeitet, wird daher – aufgrund von Beobachtungsaktivitäten zu der technischen Entwicklung von Hard- und Software generell und insbesondere der Formatentwicklung – permanent weitere Information einzupflegen sein, die wiederum zu gezielten Aktionen und Eingriffen in das System, etwa durch die Konversion einer definierten Gruppe von Objekten von einer Formatversion in eine andere, führt. Gleichzeitig unterliegt das gespeicherte Material natürlich andauernden Sicherungs- und Prüfungsläufen (bitstream preservation). Dies macht auch klar, dass der Aufwand für eine adäquate Archivierungsplattform deutlich höher ist als beim Aufbau eines „normalen“ Dokumentenservers oder institutional repository.

Diese Rahmenbedingungen legen zwingend nahe, die Langzeitverfügbarkeit digitaler Informationen, also den langfristigen Erhalt der Zugänglichkeit der Veröffentlichungen, wesentlich stärker in die existierenden Publikationsabläufe einzubinden und zu einem Teil des Publikationsprozesses zu machen. Anders als in der Vergangenheit, als sich Aufwände vor allem im eigentlichen Entstehungsprozess einer Publikation konzentrierten (und vielleicht noch in die Bereitstellung von Räumlichkeiten und Aufwänden zur Erschließung mündeten) und danach – gedacht im Nutzungskontext einer gedächtniserhaltenden Institution – nicht mehr haushaltswirksam ins Gewicht fielen, werden heute andauernde Aufwände über den kompletten Lebenszyklus (und dieser ist zumindest potentiell nicht endend) nötig sein.

Das wachsende Bewusstsein um die Notwendigkeit von Maßnahmen zur Sicherung der Langzeitverfügbarkeit digitaler Informationen einerseits und das Wissen um die Breite des Anforderungsprofils und der Anwender andererseits hat zu zwei unterschiedlichen Stoßrichtungen in den Förderaktivitäten des BMBF zur Sicherung der Langzeitverfügbarkeit digitaler Objekte geführt: Zum einen versucht nestor, unter Einbeziehung einer breiten Gruppe von Partnern aus Bibliotheken, Archiven und Museen ein Kompetenznetzwerk zum Thema Langzeitarchivierung aufzuspannen, das gleichzeitig den nationalen und internationalen Wissensaustausch anregt, aber auch durch verschiedene Aktionslinien Wissen neu generiert und umgekehrt wieder verteilt.

Gewissermaßen als Pendant im Sinne eines praxisorientierten Ansatzes fungiert demgegenüber das ebenfalls vom BMBF finanzierte Projekt „Kooperativer Aufbau eines Langzeitarchivs digitaler Informationen (kopal)“.

5. Das Projekt kopal

5.1 Grundlagen

Ziel und Zweck des durch das BMBF geförderten Projekts ist die praktische Erprobung und Implementierung eines kooperativ erstellten und betriebenen Langzeitarchivierungssystems für digitale Dokumente und Daten als eine nachnutzbare Lösung für sowohl die Langzeiterhaltung als auch die Sicherstellung der Langzeitverfügbarkeit digitaler Ressourcen. Kopal wird die technische Basis der Langzeitarchivierungsstrategie der Deutschen Nationalbibliothek bilden, der in Zukunft – im Rahmen des erweiterten Sammelauftrags – verstärkt auch die Amtlichen Publikationen umfasst. Dabei gibt es technische und operative Berührungspunkte mit den Aktivitäten der Archive.

Kopal hat aufgrund der geschilderten Rahmenbedingungen einen weit reichenden Auftrag, nämlich eine von Gedächtnisorganisationen nachnutzbare techni-

sche und organisatorische Infrastruktur zur Sicherung der Langzeitverfügbarkeit elektronischer Publikationen aufzubauen. Im Blickpunkt sind dabei auch hier nicht nur die Bibliotheken, sondern auch andere „Gedächtnisorganisationen“ wie Museen und Archive, letztlich aber jede Einrichtung, die für die dauerhafte Zugänglichkeit digitaler Objekte Sorge tragen muss.

Aufgenommen wurden die Arbeiten im Projekt zur Jahreshälfte 2004 und sind über eine Laufzeit von drei Jahren geplant. Mit einem Fördervolumen von 4,2 Mio. € (hinzu kommen noch die zum Teil erheblichen Eigenanteile der Partner) ist das Projekt auch in weltweiter Perspektive eines der größeren Vorhaben auf dem Weg zur Etablierung von Diensten zur Sicherung der Langzeitverfügbarkeit digitaler Informationen.

Ausgangspunkt des Archivsystems ist das von IBM in Zusammenarbeit mit der Nationalbibliothek der Niederlande (Koninklijke Bibliotheek, KB) entwickelte Digital Information Archiving System (DIAS).¹⁰ Die Entscheidung für dieses System fiel nach einer intensiven Marktbetrachtung und vielen Gesprächen im Jahre 2003/2004; ein wichtiger Gesichtspunkt für diese Wahl war – neben der Tatsache, dass das System auf bewährten IBM-Systemstandardkomponenten aufsetzt, die weltweit viel tausende Mal implementiert sind –, dass DIAS sich in seiner Architektur und in der Umsetzung konsequent an dem Standard „Open Archive Information System“ (OAIS) orientiert, der – seit 2003 auch über ISO etabliert – eine Art konzeptionellen Rahmen und Orientierungspunkt für Systeme abgibt, die für die Langzeitverfügbarkeit digitaler Information erstellt werden.¹¹ DIAS nimmt – darin dem OAIS-Standard folgend – die eingegliederten Daten als Submission Information Package (SIP) in einer definierten Zuordnung von erläuternden Metadaten und den eigentlichen Objekten entgegen (*ingest*). Diese werden dann im DIAS als Archival Information Package (AIP) gespeichert, in dem das Objekt (und mit ihm ggf. Tausende von Dateien) abgetrennt und in den Storage (Festplatte, Magnetband, etc.) geschrieben wird. Die Metadaten werden in einer Datenbank (Data Management) abgelegt, auf die über eine Administrationsschnittstelle zugegriffen werden kann. Bereitgestellt werden die Objekte in der Terminologie des Standards für den *access* als Dissemination Information Package (DIP).

Allerdings wies DIAS in dem in der KB eingesetzten Entwicklungsstand einige grundsätzliche Einschränkungen auf, die aus Sicht der Projektpartner unbedingt zu beheben waren:

- DIAS war als Inhouse-System entwickelt worden und kannte keinen remote access von anderen Standorten aus. Außerdem war nicht vorgesehen, dass mehrere Partner ein Archivsystem gemeinsam im Rahmen einer Kooperation betreiben. Insofern konnte das System nicht unterschiedliche Mandanten verwalten.
- DIAS verfolgte einen objektklassenbezogenen Ansatz, während die Projektpartner einen stärker einzelobjektbasierten Ansatz für notwendig erachteten. Dies schloss auch ein, dass archivrelevante Festlegungen und Zuordnungen, die einmal beim Einliefern des Objekts erfolgt waren, wieder aufhebbar bzw. umtragbar sein müssen. Dafür aber bedarf es einer objektbezogen ausgestalteten, umfangreichen Metadateninformation, die das DIAS-System so nicht kannte.
- Tools zur Homogenisierung eingestellter Objekte, die an der Einlieferungsschnittstelle ansetzen, aber auch die Software zum Zugang und zur Anzeige

¹⁰ Siehe <http://www-5.ibm.com/nl/dias/index.html> (erreichbar im Juni 2006).

¹¹ ISO Standard 2003: 14721, s. <http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf> (erreichbar im Juni 2006).

waren in DIAS von der IBM auf spezifische Klassen von Objekten (insbesondere Artikel des Verlages Elsevier) optimiert worden. Diese geschlossene Software-Lösung wurde als zu unflexibel, schlecht nachnutzbar und nicht zieladäquat eingeschätzt, stattdessen wurde der Weg offen ansprechbarer, standardisierter Schnittstellen gewählt, auf dem eine selbst entwickelte und jederzeit ausbaubare modulare Lösung aufgesetzt wurde.

5.2 Partner

Die Kooperation im Projekt kopal bilden die Bibliotheken die Deutsche Nationalbibliothek (Frankfurt am Main) und die Niedersächsische Staats- und Universitätsbibliothek (SUB Göttingen), die IBM Deutschland sowie die Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen, ein als gemeinnützige GmbH agierendes Rechenzentrum mit unterschiedlichen Kunden. Ziel dieser Verteilung von Aufgaben und Funktion war es von Anfang an, durch die dezidierte Verteilung von Zuständigkeiten und den erkennbaren Bedarf für die konsequente Abklärung der Verantwortlichkeiten, der sich aus einer solchen Projektkonstellation automatisch ergibt, gewissermaßen einen Zwang zur Kooperation und gleichzeitig zur Dokumentation der einzelnen Teilschritte in die Lösung einzubringen.

Diesen besonderen Akzent auf der Kooperation bringen auch die beiden bibliothekarischen Partner ein: Zum einen verfügen sie beide über sehr unterschiedliche Bestände an elektronischen Kollektionen (zum Beispiel DDB: Onlinedissertationen und Netzpublikationen; SUB: Digitalisate und im naturwissenschaftlich-technischen Bereich gängige Datenformate). Auf diese Weise werden sehr unterschiedliche Anforderungen in das Projekt eingebracht, die die Erfahrungspalette in der Kooperation insgesamt spürbar erweitern und die Nachnutzungsmöglichkeiten für weitere Nutzer nochmals deutlich erhöhen. Vor diesem Hintergrund arbeiten die Projektpartner eng mit dem „Kompetenznetzwerk Langzeitarchivierung“ (nestor) zusammen.

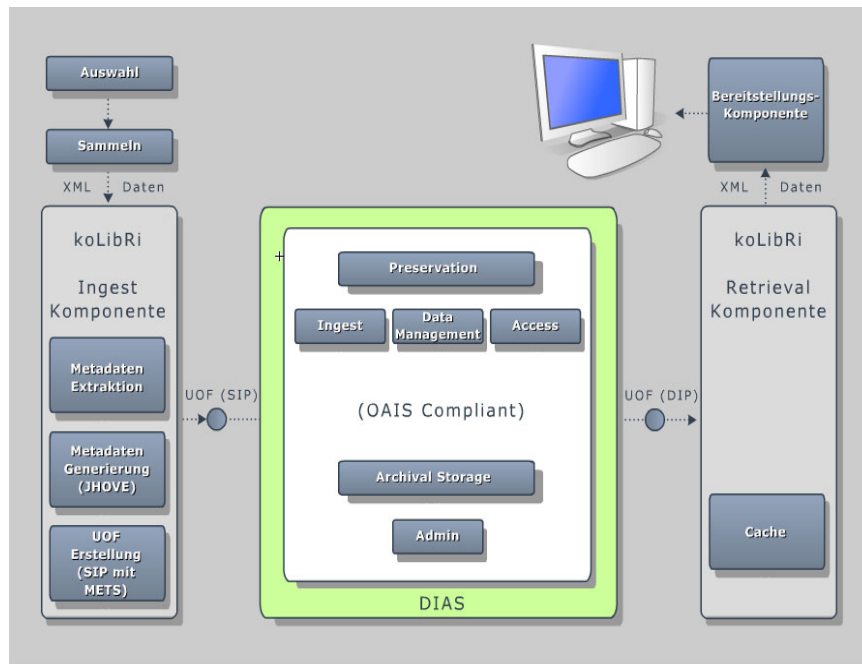
Der technische Betrieb des Langzeitarchivs ist mit der GWDG bei einem klassischen Rechenzentrum angesiedelt mit dem Ziel, vorhandene Kompetenzen in die Kooperation und das Thema einzubringen und gleichzeitig neue zu evozieren. Außerdem besteht der Bedarf, Betriebsmodelle anzuregen, zu definieren und prototypisch umzusetzen.

Ergänzt wird dieser Kreis durch den Entwicklungspartner IBM Deutschland GmbH, der über seine unmittelbaren Kontakte zu den niederländischen Partnern einerseits Knowhow-Transfer betreibt und andererseits die professionelle Anpassung der Softwarekomponenten anstrebt und dafür eine langfristig stabile Unterstützung bietet.

Im internationalen Rahmen arbeiten die Projektpartner zum einen eng mit der Königlichen Bibliothek der Niederlande (KB) zusammen, die das DIAS-Archivsystem bereits aktiv betreibt. Zusammen mit der KB werden Anforderungen an künftige Weiterentwicklungen von DIAS sowie Strategien zur Aufnahme von Preservation-Planning-Aktivitäten entwickelt. Außerdem beteiligt sich kopal an der internationalen Entwicklung und dem Austausch von Metadaten zur Langzeitarchivierung und kooperiert mit verschiedenen internationalen Projekten zur Langzeitarchivierung.

5.3 Entwicklungsziele

Eine besondere Bedeutung wurde im Projekt in der Ausweitung des in das System eingebrachten Metadatenprofils pro beschriebenes Objekt gesehen. Mit dem „Universellen Objektformat“ haben die Projektpartner ein Austausch- und Archivformat vorgelegt und implementiert, mit dem digitale Objekte zusammen mit Metadaten archiviert und zwischen Institutionen und Archivsystemen ausgetauscht werden können. Es basiert auf den Formaten METS in der Version 1.4¹² und den Langzeitarchivierungsmetadaten für elektronische Ressourcen (LMER) in der Version 1.2.¹³



Übersicht zur Software- und Systemtopologie im Projekt kopal

Einen weiteren Schwerpunkt bildet die Erstellung der Softwarekomponenten, die für die Einlieferprozeduren von Objekten verantwortlich sind. Damit wird die bereits bewährte und funktional angereicherte Kern-Software durch flexible Open Source Module erweitert, die es auch weiteren potentiellen Kooperationspartnern ermöglichen, kopal zu nutzen. Diese Tools erstellen die Deutsche Nationalbibliothek und die SUB Göttingen als auf DIAS abgestimmte Softwareprodukte: sie wurden anlässlich der CEBIT im März 2006 als „kopal Library for Retrieval and Ingest“ (koLibRI) unter einer Open Source Lizenz veröffentlicht. Einen besonderen Schwerpunkt bildet hier die Validierung der Formate eingelieferter Objekte, weil hier häufig falsche Angaben vorliegen. Hier bedient sich das Projekt einer open source Softwarelösung, die projektbasiert in den USA entstanden ist und im Kontext des Projektes kopal bereits fortentwickelt wur-

¹² Metadata Encoding & Transmission Standard; <http://www.loc.gov/standards/mets/> (letzter Zugriff 20.3.2006).

¹³ Informationen unter <http://www.ddb.de/standards/lmer/lmer.htm> (letzter Zugriff 20.3.2006).

de.¹⁴ Die entstandene koLibRI-Software steht zu Evaluierungszwecken in einer Betaversion zur Verfügung.¹⁵

6. Weiteres Arbeitsprogramm

Abgeschlossen wurden die DIAS-Basisentwicklung und -Bereitstellung sowie die DIAS-Schnittstellen-Implementierung. Darüber hinaus ist die Kernkomponente DIAS durch die IBM an die Anforderungen der Partner angepasst und als Version 2.0 ausgeliefert. Die Einrichtung der Mandantenfähigkeit ermöglicht einen flexiblen und skalierbaren Zugriff auf das System, der Fernzugriff lässt nun einen auch nicht ortsgebundenen Zugang eines Mandanten auf das System zu. Dadurch wird auch eine zukünftige Aufnahme neuer Teilnehmer in das System ermöglicht.

Damit sind zwei wesentliche Bausteine, zum einen die Bereitstellung der Software-Basiskomponenten und zum anderen die Auslieferung der Software zur Bedienung des Ingest abgeschlossen und müssen sich nun im Produktivbetrieb bewähren. Der Zugang auf die gespeicherten Objekte wird von beiden Partnern unterschiedlich angegangen: Während die Deutsche Nationalbibliothek einen mächtigen lokalen Cache aufbauen will, der im Wesentlichen nutzungsgesteuert die Abrufe beantwortet, die Deutsche Nationalbibliothek also das System eindeutig als eine Art backbone ansieht, ist die Sicht der Dinge bei den Göttinger Kollegen insofern anders, als hier von einem direkten ungepufferten Abruf aus dem System ausgegangen wird.

Angesichts des jetzt erreichten Standes sehen die nächsten Schritte im Projekt Folgendes vor:

- Optimierung der Ingest-Software für die massenhafte Einlieferung von Artikeldaten und anderen Objekten unter Performanzgesichtspunkten
- Entwicklung und Implementierung von Konzepten im Kontext der Preservation Planning Aktivitäten
- bitstream preservation: Festlegung von Rahmenbedingungen, Auslotung der technischen Möglichkeiten und Organisation des Durchgriffs in die Speicherungsprozesse pro Objekt. Die hierzu erforderlichen Speicherstrategien werden von den Projektpartnern entwickelt.
- Entwicklung und Implementierung von Migrationsprozeduren. Ein hierzu erforderliches Administrationsinterface befindet sich bereits in der Abstimmung. Darüber hinaus sollen digitale Objekte erkannt werden, die durch den technologischen Wandel bedingt in der Zukunft nicht mehr zugänglich sind. Dazu gehört auch die Anbindung einer funktionierenden internationalen Format Registry als Voraussetzung für die performante Migration großer Datenmengen.
- Entwicklung und Implementierung von Emulations-Verfahren: In diesem Themenbereich ist die internationale Kooperation ganz besonders wichtig und auch hier ist die Anbindung an Format Registries, aber auch an damit verknüpfte Tool-Registries relevant, um erfolgversprechende technische Lösungen anbieten zu können.

¹⁴ Matthias Neubauer, Thomas Wollschläger: Maschinelle Gewinnung technischer Metadaten für die Langzeitarchivierung elektronischer Publikationen. In: B.I.T. online, Zeitschrift für Bibliothek, Information und Technologie, Heft 1/2006, S. 37 – 40.

¹⁵ http://kopal.langzeitarchivierung.de/index_koLibRI.php.de (letzter Zugriff 20.3.2006).

- Entwicklung von Nachnutzungs-, Lizenzierungs- bzw. Kooperationsmodellen:
Aus der Gesamtkonstellation des Projekts sind sehr unterschiedliche Modelle für die Nachnutzung durch Partner ableitbar. Neben der Lizenzierung der kompletten Lösung sollen auch Teilmodule unabhängig voneinander genutzt werden können – ebenso dedizierte Services.
- Ergänzende Arbeiten: Ein weiterer Schwerpunkt liegt in der Dokumentation und letztlich der Zertifizierung von Verfahren und Methoden beim Betrieb eines Langzeitarchivs mit dem Ziel, hier – Stichwort „trusted repositories“-Prozesse übertragbar zu standardisieren.

Die Möglichkeit, verschiedene Mandanten mit hinsichtlich der einzutragenden Materialien ganz unterschiedlichen Selektionsprofilen in einem System zu verwalten sowie die präzise definierten Import- und Exportschnittstellen erlauben es, sehr unterschiedliche Anforderungen flexibel zu beantworten. Gleichzeitig gewährleistet die JAVA-Bibliothek koLibRI durch ihre Integrationsfähigkeit in Dokumentenmanagement- und Publikationssysteme unterschiedlicher Hersteller eine einfache und transparente Nachnutzbarkeit der entstehenden Lösung. Durch den Einsatz von bewährter IBM-Standardsoftware sind andererseits langfristige Stabilität, Performanz und Skalierbarkeit der entstandenen Lösung in hohem Maße wahrscheinlich.

Das Projekt erfüllt so den Anspruch an ein universell nutzbares Archivsystem, das technisch darauf ausgerichtet ist, durch das stabile Vorhalten der Objekte selbst und durch die (automatisierte) Erstellung eines technischen Profils zu jedem Objekt und durch die bequemen Administrationsprozesse ein Ausgangspunkt für weitere Schritte zu sein. Erst auf einer solch umfassenden Basis von Wissen zu den einzelnen eingespeicherten Objekten wird es dauerhaft möglich sein, konzeptionell und operativ Migration und Emulation konkret zu unterstützen. Ein wichtiger Pluspunkt des Systems ist auch, dass es in kopal prinzipiell keine Einschränkungen für die Art des Materials oder die Formate gibt, die in das Archiv eingespielt werden sollen. Allerdings erhöht sich, je größer die Zahl eingespielter Formate ist, auch der Aufwand zur Erweiterung der Tools, mit deren Hilfe Metadaten automatisch extrahiert werden.

7. Resümee

Am Beispiel der Amtlichen Netzpublikationen lassen sich gemeinsame Herausforderungen an Archive und Bibliotheken gut beleuchten. Begleitend zum eintretenden Wechsel von der konventionellen Druckveröffentlichung zur elektronischen Publikation wächst das Interesse daran, sich mit Lösungen zur Frage der Langzeitarchivierung und -verfügbarkeit zu beschäftigen.

Es ist deutlich geworden, dass mit dem Projekt kopal ein Ansatz vorliegt, der explizit Lösungen entwickelt, die über eine eingeschränkte bibliothekarische Sicht hinausgehen, und der weiterreichende Angebote zu offerieren in der Lage ist. Gerade im Kontext der übergreifenden Genres der Amtlichen Netzpublikationen wird deutlich, wie wichtig ein intensiver Austausch und die kooperative Nutzung von Ressourcen ist.