Maren Brodersen

Automatisiertes Abliefern von Netzpublikationen über Harvesting-Verfahren

Am 3. April 2009 veranstalteten die Deutsche Nationalbibliothek (DNB) und der Arbeitskreis Elektronisches Publizieren (AKEP) des Verlegerausschusses des Börsenvereins des Deutschen Buchhandels die dritte Informationsveranstaltung zum Thema Netzpublikationen.

Ziel der Veranstaltung war, ein Verfahren zur automatisierten Ablieferung vorzustellen, das Verlagen die Möglichkeit bietet, größere Mengen von Netzpublikationen an die DNB abzuliefern, um damit der gesetzlichen Ablieferungspflicht nachzukommen. In drei Vorträgen wurden von Mitarbeitern der DNB die folgenden Themen behandelt:

Themen

- die im vergangenen Jahr verabschiedete Pflichtablieferungsverordnung,
- die Präsentation von Netzpublikationen im Katalog der DNB sowie
- die Notwendigkeit von Verfahren zur automatisierten Ablieferung von Netzpublikationen.

Anschließend folgte ein Praxisbericht aus dem Verlagsbereich, der die Bedeutung einer Schnittstelle zum Austausch von Metadaten als Marketinginstrument darstellte, um den Datenfluss u. a. in die Bibliothekskataloge zu verkürzen und dadurch die Präsenz von E-Books zu stärken.

Ute Schwens, die Ständige Vertreterin der Generaldirektorin der DNB in Frankfurt am Main, startete die Veranstaltung mit einer kurzen Einführung zum Stand der Ablieferungsverfahren für Netzpublikationen. Dabei handelt es sich zurzeit um Webformulare für Monografien und Zeitschriften, die für kleinere Mengen von Netzpublikationen geeignet sind und damit vor allem z. B. die nicht kommerziellen oder die Selbstverlage unter den Ablieferern ansprechen. Sie betonte jedoch, dass das angestrebte Ziel die Entwicklung von automatisierten Verfahren ist und dass hier bereits mit Testpartnern gearbeitet wird. Anna Metzner-Thiele (Verlag für Standesamtswesen) begrüßte die Teilnehmerinnen und Teilnehmer im Namen des AKEP.

Zum Stand der Ablieferungsverfahren

Brigitte Wiechmann, DNB, präsentierte Details der am 23. Oktober 2008 veröffentlichten Pflichtablieferungsverordnung.[1)] Dabei ging sie vor allem auf § 7 ein, der die Beschaffenheit der Netzpublikationen und den Umfang der Ablieferungspflicht festlegt sowie § 8 und § 9, in denen die Einschränkungen der Ablieferungspflicht in Bezug auf die verschiedenen Ausgaben definiert sind. In den im Juni 2009 veröffentlichten Sammelrichtlinien werden diese Angaben näher spezifiziert. Brigitte Wiechmann betonte ebenfalls, dass für die Ablieferung größerer Mengen von Netzpublikationen automatisierte Ablieferungsverfahren notwendig sind, um den Arbeitsaufwand auf Verlags- und Bibliotheksseite möglichst gering zu halten. Erforderlich ist dabei das Einrichten einer Schnittstelle, die Bereitstellung der Metadaten und die Zugriffsmöglichkeit – ggf. auch befristet – auf die abzuliefernde Netzpublikation. Dafür sind in der Startphase Absprachen und Tests erforderlich, Ziel ist die Implementierung eines Verfahrens, das möglichst wenig Pflege und technischen Support erfordert.

Details der Pflichtablieferungsverordnung

Im Anschluss daran stellte Maren Brodersen, DNB, das Portal der DNB und hier insbesondere den Katalog als zentrale Dienstleistung vor. Auch für die Ablieferer von Netzpublikationen dient das Portal als Einstiegspunkt: Zur Registrierung und Freischaltung als Ablieferer von Netzpublikationen und nach der Freischaltung für den Zugriff auf die Webformulare.[2)] Der Schwerpunkt des Vortrages galt jedoch der Suche nach Ressourcen: Was kann mit ihnen gemacht werden, wie wird das Urheberrecht sichergestellt und welchen Nutzen hat die Verzeichnung der E-Books im Katalog für die Verlage.

Portal als Einstiegspunkt für Ablieferer

Eine Suche nach Buchtiteln kann – abhängig vom Nutzerverhalten – entweder über Google oder direkt über den Bibliothekskatalog erfolgen. Der Link führt in beiden Fällen zum Titeldatensatz. Hier sind verschiedene Informationen festgehalten: Der Hinweis auf das Format der Ressource, die Angabe von Persistent Identifiern[3)], die Anmerkung, ob die Ressource lizenzpflichtig ist, einen

Link und Informationen zum Titeldatensatz

Link zur Verlagswebsite, über die die Ressource ggf. erworben werden kann, die Gewährleistung der Langzeitarchivierung sowie ein eindeutiger Link für den Datensatz, um ihn als Bookmark speichern zu können. Autoren sind mit einem so genannten Normdatensatz der Personennamendatei[4] verlinkt und dieser führt ggf. zu weiteren Publikationen. Bei Autoren, die in Wikipedia zu finden sind, wird ebenfalls auf den Normdatensatz im Katalog verlinkt.[5]

Zugriffsrechte

Für das abgelieferte Archivexemplar gelten die Rechte, die der Ablieferer der DNB einräumt. Ist ein Originalexemplar lizenzpflichtig, dann ist in der Regel der Zugriff auf das Archivexemplar nur in den Lesesälen möglich. Die Leseplätze haben keinen Internetanschluss und keine USB-Schnittstelle. Ein download ist somit nicht möglich. Darüber hinaus kann ein Ablieferer aber die Zugriffsrechte auf registrierte Benutzer – mit gültigem Benutzerausweis – ausweiten, die dann auch außerhalb des Lesesaals zugreifen können. Abgeliefert werden muss lediglich eine Ausgabe der Netzpublikation, aber der Zugriff wird sowohl in Frankfurt am Main als auch in Leipzig ermöglicht.

Diskussion

In der folgenden Diskussion ging es insbesondere um die Software, die E-Book-Reader und die Vielfalt der Angebote, die sowohl für die Verlage als auch für die Bibliotheken ein Problem darstellen. Dies gilt besonders dann, wenn die Netzpublikationen von einer bestimmten Hardware abhängig sind und sich daraus auch Probleme für die Ablieferung ergeben. Diskutiert wurde auch über XML als das Format für viele Endgeräte sowie das speziell für E-Books entwickelte epub-Format.[6] Es wurde jedoch in Frage gestellt, ob dies ein zukunftsträchtiger Standard ist, denn problematisiert wurden insbesondere die Abweichungen, die sich leider häufig auch bei Standards einschleichen. Die Angabe der URL der Produkt-Website im Katalog wurde durchaus positiv gesehen.

Harvesting-Verfahren

Jürgen Kett, DNB, führte in das Thema der automatisierten Ablieferung über Harvesting-Verfahren ein, indem er kurz das Problem der Ablieferung für die DNB darstellte. Einer Vielzahl von Partnern stehen begrenzte Ressourcen sowohl auf Verlags- als auch auf der Bibliotheksseite gegenüber, die sich mit ganz unterschiedlichen organisatorischen und technischen Bedingungen auseinandersetzen müssen – in der Realität existieren unterschiedliche Workflows, Formate, Protokolle etc. Daher ist es erforderlich, dass die möglichen Ablieferungsverfahren überschaubar und realisierbar bleiben.

Protokoll OAI-PMH

Das erste bereits implementierte Verfahren basiert auf dem technischen Protokoll OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting). Die Open Archive Initiative bietet eine offene Schnittstelle, die zum Austausch von XML-Metadaten geeignet ist. Dabei stellt ein Repository bzw. Server die Metadaten bereit und ein Harvester bzw. Client sammelt diese Daten ein. Abgeholt werden nur die Metadaten, die eine so genannte Transfer- oder Access-URL enthalten, über die der Zugriff auf die Netzpublikation ermöglicht wird. Dazu wird ein http-basiertes Protokoll (OAI-PMH) eingesetzt und die Anfragen erfolgen über einfache GET- oder POST-Requests. Der Vorteil besteht darin, dass das Verfahren durch Programmiersprachen unterstützt und eine bewährte Technik angewandt wird.[7] Die Kommunikation oder der Austausch erfolgen über Fragen und Antworten mittels einer gesicherten Verbindung. Die DNB gewährleistet dabei diese Absicherung der Kommunikation und es kommen Standardverfahren wie z. B. SSL oder IP-Beschränkungen zum Einsatz.

Die Realisierung der automatischen Ablieferung wird für jeden Verlag im Rahmen eines gesonderten Projekts realisiert, das beide Seiten aktiv als Projektpartner durchführen. Erst im Rahmen dieser Projekte werden die technischen Details der Realisierung festgelegt und eine Entscheidung über das einzusetzende Ablieferungsverfahren gefällt (es sind neben OAI prinzipiell auch andere Lieferverfahren möglich).

Diskutiert wurde vor allem über den Zugriff auf die Netzpublikationen bzw. den Prozess der Abholung. Vorgeschlagen wurde u. a. ein temporärer Zugriff auf das abzuliefernde E-Book. Dies ist über vereinbarte Zeiträume für die Abholung über eine OAI-Schnittstelle durchaus möglich. Thematisiert wurden auch verschiedene Formen der Benachrichtigung, die die Abholung bestätigen: Eine E-Mail nach Abholung oder das Löschen der abgeholten Sätze.

Die Ablieferung von Metadaten über eine OAI-Schnittstelle wird zurzeit mit dem Springer Verlag umgesetzt. Dirk Fernholz von Springer Media berichtete über seine Erfahrungen mit der E-Book-

Ablieferung von Mediadaten über OAI

Nutzung und den Vorteilen der automatisierten Ablieferung für das Marketing. Er wies auf die unproblematische Ablieferung über die OAI-Schnittstelle hin. Nach seinen Erfahrungen sucht die Zielgruppe der Springer E-Books aus dem STM-Bereich (Science, Technology, Medicine) hauptsächlich in den Katalogen der Bibliotheken nach relevanter Literatur und soll dort neben den Printpublikationen auch die E-Books finden. An Beispielen von australischen und neuseeländischen Universitätsbibliotheken belegte er, dass durch den Nachweis in den Katalogen die Zugriffe auf die Volltexte ansteigen. Auch am Beispiel der Hochschule für Technik in Stuttgart war sichtbar, dass die Zugriffe für die E-Books, die im Katalog verzeichnet sind, eine steigende Tendenz haben. Der Zugriff erfolgt hier über den Link im Katalogsatz direkt auf den Publikationsserver von Springer.

Erhöhte Relevanz von Metadaten durch veränderte Produktionsprozesse

Eine Ursache für die erhöhte Relevanz von Metadaten sind die veränderten Produktionsprozesse. Musste das gedruckte Buch ursprünglich den gesamten Herstellungsprozess durchlaufen, bevor es in die Bibliothek und damit in den Katalog gelangte, so wurden mit der Umstellung auf digitale Produktionsverfahren die Metadaten relativ früh im Publikationsprozess erstellt und stehen damit früher zur Weitergabe zur Verfügung. Mit der Erstellung der Volltexte im XML-Format als Basistext werden die Metadaten bereits am Anfang des Produktionsprozesses erstellt und können in verschiedenen Formaten angeboten und ausgeliefert werden. Dies verkürzt und verändert die Datenflüsse in die unterschiedlichen Datenbanken. Bislang werden über das Verzeichnis lieferbarer Bücher (VLB) nur gedruckte Bücher angekündigt, aber auch hier ist damit zu rechnen, dass die Neuerungen den Datenaustausch verändern werden. Die Realisierung einer OAI-Schnittstelle kann dabei ein Weg zur Verbreitung der Metadaten sein, denn darüber kann beispielsweise auch die Lieferung der aufbereiteten Daten der DNB zurück an den Verlag erfolgen.

Identifier für E-Books

Außerdem wurde u. a. über den Zweck der ISBN für E-Books diskutiert. Ein Identifier ist notwendig, aber hier bieten sich Alternativen wie z. B. DOI oder URN, also spezifisch elektronische Identifier.

Alternative Ablieferungsmodelle wurden nicht diskutiert, aber es wäre erfreulich, wenn sich interessierte Verlage melden würden, die sich in diesem Bereich Alternativen vorstellen können.

Anmerkungen

1 Pflichtablieferungsverordnung (PflAV) <http://bundesrecht.juris.de/pflav/index.html>

2 Sammelrichtlinien <http://files.d-nb.de/pdf/sammelrichtlinien.pdf>

3 Informationen zu den Ablieferungsverfahren finden sich unter: <http://www.d-nb.de/netzpub/ablief/np_ablief.htm>

4 Jede abgelieferte Netzpublikation erhält ein URN. Die Vergabe erfolgt automatisch. Informationen zum URN unter: <http://www.persistent-identifier.de/>

5 Personennamendatei (PND) <http://www.d-nb.de/standardisierung/normdateien/pnd.htm>

6 Beispiel: Der Autor Bertolt Brecht <http://d-nb.info/gnd/118514768>

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

8 Hinweis zur automatisierten Ablieferung über Harvesting-Verfahren unter: <http://www.d-nb.de/netzpub/ablief/pdf/automatisierte_ablieferung.pdf>
Einstiegsinformation der Open Archive Initiative zum Protokoll unter: <http://www.openarchives.org/OAI/openarchivesprotocol.htm>
»Guidelines for Repository Implementers« unter: <http://www.openarchives.org/OAI/2.0/guidelines-repository.htm>