

Vortrag auf dem Bibliothekartag in Hamburg, 24. Mai 2012, TK 9: Metadatenbaukasten
Maren Brodersen, Informationstechnik, Deutsche Nationalbibliothek

XMetaDissPlus: Publikationstypen und die Herausforderung von Zeitschriftenlieferungen

Das Metadatenformat XMetaDissPlus hat im Laufe seiner nunmehr 14-jährigen Nutzung mehrere Entwicklungsstufen durchlaufen, die teilweise technisch bedingt waren, teilweise aber auch auf den veränderten Sammlungsauftrag der Deutschen Nationalbibliothek (DNB) zurückzuführen sind. So ergaben sich mit der Anwendung des Protokolls OAI-PMH und damit der Möglichkeit, Metadaten automatisiert harvesten zu können, andere Anforderungen sowohl an den Standard als auch an den Workflow.

Anfangs war der Standard noch stark an das Ablieferungsverfahren für Online-Dissertationen ausgerichtet. Inzwischen wurde er erweitert, damit ist XMetaDissPlus auch für andere Publikationstypen nutzbar. Für die automatisierte Ablieferung stellen die Zeitschriftenlieferungen dabei eine besondere Herausforderung dar.

Von MetaDiss ...

Seit Juli 1998 sammelt die DNB Online-Dissertationen und –Habilitationen und betreut das Metadatenformat für Online-Hochschulschriften. Der erste Standard, der dafür entwickelt wurde, nannte sich MetaDiss, war HTML basiert und verwendete Dublin Core sowie eigens dafür erstellte DDB-Elemente¹. Als Schnittstellen für die Übertragung der Metadaten wurde ein Webformular und die Ablieferung per E-Mail angeboten. Die Verarbeitung intern erfolgte mit Hilfe einer Software, die einen Rumpfdatensatz anlegte, der manuell angereichert wurde. Auch die Archivierung des Objektes erfolgte zunächst halbautomatisiert.

... zu XMetaDiss ...

Mit der Möglichkeit des Metadaten Harvesting durch das OAI-Protokoll (OAI-PMH) wurde eine Anpassung des Standards erforderlich. Das Harvesting von Metadaten setzt die Übertragung von XML-Daten voraus, da diese regelbasiert hierarchisch strukturiert sind und somit u.a. eine korrekte automatische Auswertung ermöglichen. Als Grundlage der im Format aufgeführten Datenelemente war neben dem Dublin-Core-Metadaten-Element-Set auch hier die Einbindung eigener ddb-Elemente² erforderlich. Erweitert wurde der Standard zusätzlich um das Metadatenset der Networked Digital Library for Theses and Dissertations ETD-MS (Electronic Theses and Dissertations Metadata Set), um Metadaten zur Langzeitarchivierung auf Basis der New Zealands National Library Preservation Metadata sowie um ein Set zur Erfassung von personenbezogenen Daten (XMetaPers).

Integriert wurde auch das Element „Transfer-URL“, um über die gelieferte URL eine Abholung der Publikation zu ermöglichen. Zu diesem Zeitpunkt war es aber kein Pflichtfeld und wurde nur für

¹ DDB ist ein Prefix und in diesem Fall historisch bedingt. Es steht für die vor Juni 2006 verwendete Bezeichnung der Deutschen Nationalbibliothek als „Die Deutsche Bibliothek“.

² Siehe oben, allerdings in diesem Fall in Kleinbeschreibung zur Einhaltung der XML-Konventionen.

Mehrdateiendokumente genutzt. Desweiteren wurde die URN als Persistent Identifier als Pflichtfeld eingeführt. Der Standard war auch weiterhin nur für die Ablieferung von Online-Dissertationen und -Habilitationen gedacht. Neben der Ablieferung per E-Mail und über das Webformular konnten die Metadaten jetzt auch geharvestet werden, aber der Workflow blieb halbautomatisch und arbeitsintensiv.

... und weiter zu XMetaDissPlus

Mit Inkrafttreten des Gesetzes über die Deutsche Nationalbibliothek (DNBG) vom 22. Juni 2006 erweiterte sich für die DNB der Auftrag der Sammlung, Erschließung, Verzeichnung und Archivierung auch auf unkörperliche Medienwerke (Netzpublikationen). Damit waren alle Online-Hochschulschriften, die der Öffentlichkeit zugänglich gemacht werden, ablieferungspflichtig und nicht nur die Hochschulprüfungsarbeiten. Um den Universitätsbibliotheken die Möglichkeit zu bieten, XMetaDiss auch weiterhin für die OAI-Schnittstelle nutzen zu können, wurde die Erweiterung auf andere Publikationstypen in Zusammenarbeit mit dem Bibliotheksservice-Zentrum Baden-Württemberg (BSZ) geplant, die bereits für die internen Abläufe in XMetaDiss eine Publikationstypenliste integriert hatten und diese Version XMetaDissPlus nannten. Das Ziel war jedoch, für die zukünftigen Publikationstypen auf ein standardisiertes Vokabular zu setzen. In einer Unterarbeitsgruppe der AG Elektronisches Publizieren der DINI e.V. wurde daraufhin ein »Gemeinsames Vokabular der Publikations- und Dokumenttypen« erarbeitet. Darauf aufbauend wurden in enger Zusammenarbeit zwischen dem BSZ und der DNB die Referenzbeschreibung, die Schemata und Beispiele für XMetaDissPlus erstellt.

Voraussetzung für die automatisierte Ablieferung

Die Erweiterung von XMetaDiss auf XMetaDissPlus wurde notwendig durch die Umstellung auf die automatisierte Ablieferung der Netzpublikationen, die bereits seit Oktober 2009 mit dem Metadatenstandard ONIX for Books 2.1 von Verlagen genutzt wird. Damit waren für die abliefernden Universitätsbibliotheken jedoch einige Veränderungen im Workflow verbunden. Wurden die importierten Datensätze bis zur Umstellung noch manuell nachgebessert, so mussten die Metadaten im automatisierten Verfahren valides XML liefern, schemakonform sein und korrekte Inhalte in den Elementen übermitteln. Außerdem wurde die Transfer-URL jetzt für die automatisierte Abholung der Publikation im Ablieferungsprozess genutzt und ist inzwischen ein Pflichtfeld.

Die Umstellung erfolgte in Stufen. In vielen Fällen wurde die automatisierte Ablieferung zuerst für die Online-Dissertationen umgesetzt, um das Verfahren zu implementieren. Die Einführung der Publikationstypen zog bei den meisten Universitätsbibliotheken ein Mapping der bisher verwendeten Begriffe auf das neue Vokabular nach sich.

Anpassungen für Zeitschriften in XMetaDissPlus

Die Integration der weiteren Publikationstypen erforderte auch eine Anpassung der Elemente, die bislang nur auf Online-Dissertationen ausgerichtet waren. Teilweise wurden Pflichtfelder entfernt, teilweise wurden Elemente um Attribute erweitert. Dies war vor allem im Rahmen der Erweiterung auf die automatisierte Ablieferung von Zeitschriftenlieferungen (Artikel und Hefte) und die damit verbundene automatisierte Verknüpfung mit den zugehörigen Zeitschriftentiteln der ZDB notwendig.

Dafür wurden folgende Publikationstypen eingeführt:

article, *contributionToPeriodical* und *PeriodicalPart* für die Zeitschriftenlieferungen und *Periodical* für den Zeitschriftentitel. In dem Element 29 „Hochschulschrift ist Teil von“ wurden die folgenden Attribute für die Identifier eingeführt, die die Verknüpfung ermöglichen sollten:

ddb:DNB-ZStitelID → ID des DNB-Titelsatzes

ddb:Erstkat-ID → für die ZDB-ID

ddb:ZStitelID → systemeigener Identifier oder wahlweise auch ISSN, URN oder DOI

Außerdem wurden Attribute für die strukturierte Angabe von Ausgabeninformationen eingeführt:

ddb:ZS-Volume – für die Angabe von Band, Volume oder Jahrgang

ddb:ZS-Issue – für die Angabe von Heft oder Nummer

ddb:ZS-Ausgabe – für nicht normierte Angaben wie z. B. Stand- oder Datumsangaben

Die technische Umsetzung für das Verfahren ergab sich aus der Ablieferung über das Webformular. Hier wird bereits in jeden Zeitschriftentitelsatz, für den ein Ablieferer Hefte oder Artikel abliefern, die ID des jeweiligen Ablieferers eingetragen. Voraussetzung ist, dass dieser Satz bereits in die Zeitschriftendatenbank (ZDB) integriert wurde. Ist dies nicht der Fall, dann erfolgt in diesem Zuge auch die Aufnahme der Zeitschriftentitel in die ZDB. Im Webformular wird dem Ablieferer eine Liste seiner Zeitschriftentitel angezeigt. Hier muss für die Ablieferung eine Auswahl getroffen werden und diesem Titel werden dann entsprechend die abgelieferten Artikel oder Hefte zugeordnet. Ausgewählt werden muss auch, ob es sich um die Ablieferung eines Artikels oder eines Heftes handelt.

The image shows a web form titled "Angaben zur Netzpublikation" (Details for Network Publication). The form is divided into sections. The first section is "Allgemeine Angaben" (General Information), which includes a radio button selection for "Art des abgelieferten Dokuments:" (Type of delivered document) with options "Artikel" (Article) and "Heft" (Issue). Below this is a dropdown menu for "Titel der Zeitschrift:" (Journal title) with the placeholder text "Bitte wählen:" (Please select:). The next section is "Eigenständiger Titel des Artikels/Heftes (* für Artikel):" (Independent title of the article/issue (* for article)), followed by a text input field for "Erscheinungsjahr online (JJJJ):" (Year of online appearance (YYYY)). Below that is a text input field for "Adresse der Netzpublikation (URL):" (Network publication address (URL)). The next section is "Ausgabe" (Issue), which includes a text input field for "Band/Volume/Jahrgang (keine Jahresangabe):" (Volume/Year (no year specification)), a text input field for "Heft/Nummer:" (Issue/Number), a text input field for "Berichtsjahr (JJJJ oder JJJJ-JJJJ):" (Report year (YYYY or YYYY-YYYY)), a text input field for "Stand/Version/andere Angaben:" (Status/Version/other information), and a text input field for "Seitenangabe (von-bis):" (Page specification (from-to)).

Abb. 1: Ansicht des Webformulars für die Ablieferung von Artikeln und Heften

Für die automatisierte Ablieferung wurde ein weiteres Feld im Titelsatz integriert, in dem ein Identifier eingetragen wird, der in den Metadaten mitgeliefert werden muss, die oben genannte *DNB-ZStitelID*, *ZStitelID* oder die *Erstkat-ID*.

Vor der Ablieferung muss mit dem Ablieferer geklärt werden, für welche Zeitschriftentitel er Artikel oder Hefte abliefern wird und welchen Identifier er mitliefern kann. Da nicht alle Zeitschriftentitel über eine ISSN oder eine URN verfügen, wird gemeinsam mit dem Ablieferer nach einer Lösung gesucht. Wird im eigenen Katalog eine ID für die Verknüpfung verwendet, so kann diese auch für die automatisierte Verknüpfung genutzt werden. Der systemeigene Identifier kann eine Zahl, eine Abkürzung des Zeitschriftentitels oder eine Ziffern- und Buchstabenkombination sein. Durch die individuelle ID des jeweiligen Ablieferers ist die Kombination in jedem Fall eindeutig, wie auch die folgende Katalogansicht zeigt.

```

2005 1868-6869*Africa spectrum (Internet)
2010 1868-6869*
2013 |p|0002-0397*
2105 09.A48.0229
2110 2466495-9
2115 995808430 --> ID des DNB bzw. ZDB-Titelsatzes
2150 1868-6869 --> systemeigener Identifier (ISSN, URN, ...)
2198 25165824 --> ID des Ablieferers
2199 eb50905078
2240 ZDB:2466495-9 --> ZDB-ID
2242 |o|316257973
2242 |||2009235670
2242 |o|474838682
3120 !1014711959!Institut für Afrika-Studien$gHamburg
3260 Africa-Spectrum
4000 Africa Spectrum [(Elektronische Ressource)] / GIGA German Institute of Global and Area Studies, Institute of African Affairs,
4024 /b1966-
4025 1.1966 -
4030 New York, NY : JSTOR
4030 Hamburg : Hamburg Univ. Press
4030 Hamburg : Inst. of African Affairs
4060 Online-Ressource
4083 =A http://d-nb.info/991346203

```

Abb. 2: Ansicht des Zeitschriftentitels in der internen Datenbank

Die Verknüpfung zwischen Zeitschriftentitel und Artikel bzw. Heften ist in der Katalogansicht sichtbar.

Link zu diesem Datensatz	http://d-nb.info/991346203
Zeitschrift/Serie	Africa Spectrum [Elektronische Ressource] / GIGA German Institute of Global and Area Studies, Institute of African Affairs, ...
Organisation(en)	Institut für Afrika-Studien <Hamburg>
Verleger	New York, NY : JSTOR Hamburg : Hamburg Univ. Press Hamburg : Inst. of African Affairs
Erscheinungsverlauf	1.1966 -
Umfang/Format	Online-Ressource
Parallele Ausgabe(n)	Druckausg.: Africa-Spectrum
Anmerkungen	Nebent.: Afrika-Spectrum Gesehen am 13.08.09 Ersch. unregelmäßig Text teils dt., teils engl., teils franz. Langzeitarchivierung gewährleistet
URL	http://d-nb.info/991346203 http://www.bibliothek.uni-regensburg.de/... http://www.giga-hamburg.de/index.php?fil... http://hup.sub.uni-hamburg.de/giga/afsp http://www.jstor.org/action/showPublicat... http://search.ebscohost.com/kostenfrei kostenfrei
ISSN	Autorisierte ISSN: 1868-6869 Africa spectrum (Internet) ISSN der Vorlage: 1868-6869 ISSN der parallelen gedruckten Ausgabe: 0002-0397
Sachgruppe(n)	320 Politik
Zugehörige Bände	32 Publikationen 1. Review: Robert H. Bates: When Things Fell Apart. State Failure in Late-Century Africa (2008) [Elektronische Ressource] In: Africa Spectrum Bd. 44, Nr. 3, Erstausgabe: 165-168

Abb. 3: Ansicht des Titels Africa Spectrum <http://d-nb.info/991346203> im Katalog der DNB

Die Umsetzung eines XMetaDissPlus „record“ für Artikel sieht (hier in Auszügen) wie folgt aus:

```
<xMetaDiss:xMetaDiss xmlns:xMetaDiss=http://www.d-nb.de/standards/xmetadissplus/
...
<dc:title xsi:type="ddb:titleISO639-2" lang="eng">Ethiopian Nationalism: An Ideology to
Transcend All Odds</dc:title>
<!-- Bei Artikeln sollte die Angaben der Autoren erfolgen -->
<dc:creator xsi:type="pc:MetaPers">
<pc:person>
<pc:name type="nameUsedByThePerson">
<pc:foreName>Belachew</pc:foreName>
<pc:surName>Gebrewold</pc:surName>
</pc:name>
</pc:person>
</dc:creator>
...
<dcterms:issued xsi:type="dcterms:W3CDTF">2009-05-27</dcterms:issued>
<dc:type xsi:type="dini:PublType">article</dc:type>
<dc:identifier xsi:type="urn:nbn">urn:nbn:de:gbv:18-4-308</dc:identifier>
<dcterms:medium xsi:type="dcterms:IMT">application/pdf</dcterms:medium>
<dc:language xsi:type="dcterms:ISO639-2">eng</dc:language>
<dcterms:isPartOf xsi:type="ddb:ZSTitelID">1868-6869</dcterms:isPartOf>
<dcterms:isPartOf xsi:type="ddb:ZS-Volume">44</dcterms:isPartOf>
<dcterms:isPartOf xsi:type="ddb:ZS-Issue">1</dcterms:isPartOf>
...
</xMetaDiss:xMetaDiss>
```

Abbildung 4 zeigt die Ansicht eines Artikelsatzes, der mit dem Titelsatz verknüpft ist. Für die Ablieferung eines Artikels sind nur wenige Metadaten erforderlich. Dazu gehört neben Titel und Autor nur das Erscheinungsjahr, die Angaben zur Ausgabe und der Identifier für die Verknüpfung.

	
Link zu diesem Datensatz	http://d-nb.info/995932042
Person(en)	Gebrewold, Belachew
Titel	Ethiopian Nationalism: An Ideology to Transcend All Odds [Elektronische Ressource] / Belachew Gebrewold
Persistent Identifier	urn:nbn:de:gbv:18-4-308
URL	Archivserver der Deutschen Nationalbibliothek http://hup.sub.uni-hamburg.de/giga/afsp/...
In:	Africa Spectrum, Bd. 44, Nr. 1: 79-97
Erscheinungsjahr	2009
Sachgruppe(n)	320 Politik

Abb. 4: Ansicht eines Artikels zum oben genannten Titelsatz <http://d-nb.info/995932042>

XMetaDissPlus: wie geht es weiter?

Für die automatisierte Zeitschriftenablieferung wurde eine gute Lösung gefunden, aber in der täglichen Arbeit zeigt sich, dass weitere Anpassungen folgen werden. Aus der Community der Nutzer kommt eine wachsende Zahl an Anforderungen. Meist sind es nur kleine Änderungen und

Anpassungen an den Elementen, die gewünscht werden. Häufig betreffen die Anforderungen und Anfragen Probleme bezüglich der URN-Vergabe und damit verbunden die Praxis, wie mit Versionsänderungen umgegangen wird oder umgegangen werden sollte.

Eine Anforderung, die sicherlich umfassendere Veränderungen nach sich zieht, ist die Nutzung von XMetaDissPlus auch für Print-Publikationen. Hier wird die Einführung mindestens eines weiteren Elementes nötig und es stellt sich auch die Frage, ob es sich dabei um ein Pflichtfeld handeln soll.

Diese und andere Fragen zum Metadatenstandard XMetaDissPlus sollen deshalb am 20. Juni 2012 in Frankfurt in einem Workshop diskutiert werden. Daraus wird sich dann ergeben, in welche Richtung sich der Standard weiterentwickeln wird.

Für weitere Fragen stehe ich gern zur Verfügung.

Maren Brodersen

m.brodersen@dnb.de

Weitere Informationen

Überblicksseite XMetaDissPlus:

<http://www.dnb.de/DE/Standardisierung/Metadaten/xMetadissPlus.html>

Referenzbeschreibung XMetaDissPlus 2.2:

<http://nbn-resolving.de/urn:nbn:de:101-2012022107>

Gemeinsames Vokabular Publikations- und Dokumenttypen:

<http://nbn-resolving.de/urn:nbn:de:kobv:11-100109998>

Kernset für die Ablieferung an die Deutsche Nationalbibliothek:

<http://nbn-resolving.de/urn:nbn:de:101-2012022202>