

Christine Frodl, Nicole Germann, Sarah Hartmann, Ute Schwens, Brigitte Wiechmann

Informationsveranstaltung zum Thema Metadaten

Gemeinsamer Workshop der Deutschen Nationalbibliothek (DNB) und des Arbeitskreises Elektronisches Publizieren (AKEP)

Gemeinsame
Begriffsdefinition
als Ziel

Am 22. Februar 2008 veranstalteten der Arbeitskreis Elektronisches Publizieren des Verlegerausschusses des Börsenvereins des Deutschen Buchhandels (AKEP) und die DNB in Zusammenarbeit mit der Marketing- und Verlagsservice des Buchhandels GmbH (MVB) einen Workshop zum Thema Metadaten. Ziel des Workshops, an dem 80 Vertreter aus Verlagen, Bibliotheken und Dienstleistungsunternehmen teilnahmen, war es, zu einer gemeinsamen Definition des Begriffs »Metadaten« und zu gemeinsamen Festlegungen bezüglich Metadaten für Netzpublikationen zu kommen, die so praktikabel und homogen wie möglich sein sollten, um eventuelle Doppelarbeiten zu vermeiden.

Da das Thema Metadaten und Netzpublikationen sehr umfangreich ist, lag der Schwerpunkt der Veranstaltung auf monografischen, abgeschlossenen Netzpublikationen.

Schwerpunkt
monografische
Netzpublikationen

Der Einführungsvortrag von Christine Frodl (DNB) hatte zum Ziel, sich dem komplexen Thema Metadaten anzunähern und darzustellen, welche Bedeutung Metadaten haben und welche Rolle sie in verlegerischen und bibliothekarischen Prozessen spielen. Verlage, der Buchhandel und die Bibliotheken beschäftigen sich alle mit Ressourcen, somit gibt es vielfache inhaltliche Überschneidungen bei Beschreibungskonzepten und Metadatenstandards. Was genau ist unter Metadaten zu verstehen? Metadaten sind strukturierte (und idealerweise standardisierte) maschinenlesbare Daten über Inhalt und Form (digitaler) Ressourcen oder Objekte, die der Identifizierung, Erschließung und dem Wiederauffinden dieser Ressourcen dienen. Metadaten entstehen in Verlagen bei der Manuskripterstellung, in Content-Management-Systemen, sie sind wichtig beim Marketing der Verlags-

produkte sowie im Verkaufsprozess. Kunden von gedruckten und elektronischen Publikationen greifen mithilfe von Metadaten, wie z. B. Titel, Autor oder ISBN auf die Verlagsprodukte zu. Autoren zitieren Werke und tun dies mit bibliografischen Angaben, und wenn Bibliotheken Bücher für ihre Lehrbuchsammlung kaufen, wählen sie die Titel überwiegend anhand der Metadaten aus.

Netzpublikationen zeichnen sich gegenüber traditionellen Printmedien vor allem durch ihre Heterogenität, ihre Granularität, ihre Dynamik und ihre Instabilität aus. Sie sind häufig »offen« konzipiert und nicht immer ist inhaltliche Qualität gewährleistet. Oft werden Inhalte auch an andere Orte verlagert, vernetzt und neu zusammengesetzt. Das bedeutet, dass bisher verwendete bibliografische Standards und Methoden aktualisiert und so den modernen Anforderungen angepasst werden müssen. Um eine Interoperabilität beim Datenaustausch zu erreichen, müssen standardisierte Übereinkünfte zwischen den am Metadaten austausch beteiligten Parteien getroffen werden. Unterschiedliche Anwendergruppen im Publikationsbereich nutzen unterschiedliche Metadatenformate. Näher vorgestellt wurden die Standards MARC 21, ONIX und Dublin Core, da diese häufig im Verlags- und Bibliotheksbereich Anwendung finden.

Im vergangenen Jahr wurde in der DNB ein »Metadaten-Kernset« entwickelt. Es stellt die Definition eines minimalen Kernsets an Metadatenelementen für die Anlieferung von Metadaten elektronischer Ressourcen an die DNB dar (also den Mindestumfang an Metadaten). Die Menge der Metadatenelemente, die über die OAI-Schnittstelle geliefert werden, muss jedoch nicht auf das Kernset beschränkt sein. Bei der Erstellung des Kernsets wurde von der Fragestellung ausgegangen, welche Metadatenelemente die DNB unbedingt und mindestens benötigt, um Netzpublikationen den Anforderungen entsprechend bearbeiten zu können. Diese Elemente wurden dann in einem ersten Schritt auf das ONIX-Format abgebildet, da davon ausgegangen

Heterogen,
granular,
dynamisch,
instabil

Metadaten-
Kernset
als Grundlage

wurde, dass Verlage ihre Metadaten im ONIX-Format bereithalten. Die vorliegende Ausgabe umfasst zunächst so genannte »monografische« elektronische Publikationen. Es ist geplant, dieses »Metadaten-Kernset« um weitere Elemente für weitere Publikationstypen zu ergänzen, z. B. um Musikalien und Musikträger, periodische Publikationen, kartografisches Material. In Arbeit ist auch die Berücksichtigung weiterer Metadatenformate, wie z. B. MARC 21.

Metadaten
im Verlag

Nach der Einführung in die Thematik durch Christine Frodl gab Christian Kohl, Project Manager Electronic Publishing bei Walter de Gruyter GmbH & Co KG, einen Überblick über die Erstellung und Verwaltung von Metadaten im Verlag Walter de Gruyter. Die Verwaltung und Pflege der Metadaten bei de Gruyter erfolgt an unterschiedlichen Stellen bzw. in verschiedenen Abteilungen im Verlag: Standardmetadaten, die für alle Produkte vorgehalten werden, werden mit einer Verlagssoftware, produktspezifische Metadaten hingegen werden im Content-Management-System gepflegt. Die Metadaten werden vom Lektorat erstellt und zu einem späteren Zeitpunkt des Geschäftsprozesses von Mitarbeitern der Herstellung ergänzt. Deskriptive Metadaten, wie z. B. Titel, Untertitel, englische Titel, Stichwörter, Vorschautexte, Informationen über Autoren werden in der Regel standardmäßig zu jedem Titel erfasst. Zusätzlich werden produktspezifische Metadaten, wie beispielsweise bestimmte Sortierschlüssel oder Klassifikationen, und administrative Metadaten, die für das Controlling, die Herstellung und den Vertrieb benötigt werden, verwaltet. Am Beispiel des Onlineproduktes »Pschyrembel Klinisches Wörterbuch Online« präsentierte Christian Kohl, welche Metadaten erfasst wurden und wie diese für die Onlinepräsentation des Produktes genutzt werden. Die Metadaten werden sowohl für verlagsinterne Prozesse benötigt als auch an externe Partner weitergegeben. In diesem Zusammenhang wurde betont, wie wichtig die Produktion von korrekten Metadaten ist, die mit anderen Partnern ausgetauscht werden, da die Metadaten den Wert einer Netzpublikation erhöhen.

Im Anschluss an den Vortrag von Christian Kohl berichtete Martin Lüning vom Verzeichnis Lieferbarer Bücher (VLB) über die Erfahrungen mit

ONIX-Titelmeldungen. An das VLB wird bisher nur eine geringe Anzahl an Netzpublikationen gemeldet, wobei sich diese Meldungen kaum von den Titelmeldungen für Printprodukte unterscheiden. Die Anwendungsfälle wurden daher vor allem aus Sicht des Standardformats »ONIX for Books« (Version 2.1) dargestellt. Außer dem Standard ONIX for Books in der Version 2.1, rev. 03, zu dem umfangreiche Code-Listen bereits in der 8. Ausgabe vorliegen, existieren weitere ONIX-Formate. ONIX DOI Name Registration Formats, Version 1.1 umfasst vier Teilformate für Monografien und sechs Teilformate für Zeitschriften. ONIX for Serials beinhaltet vier unterschiedliche Formate, wovon drei aber noch nicht freigegeben sind. Das Format ONIX for Licensing Terms ist bisher ebenfalls noch nicht freigegeben. Bislang gibt es leider keine Best-Practice-Lösung für die Beschreibung von Netzpublikationen. Gemäß dem Metadaten-Kernset der DNB müssen für die ONIX-Titelmeldung zwei Typen von URLs übermittelt werden: Die URL der Produktwebsite/Frontpage und/oder die URL, die den direkten Zugriff auf die Ressource erlaubt. Eine der beiden URLs ist für eine Titelmeldung aber unbedingt erforderlich. In ONIX for Books können Links auf Produktwebsites spezifiziert werden, ONIX-DOI spezifiziert Transfer-URLs.

ONIX for Books bietet zurzeit keinen Container für Lizenzmetadaten. Es ist bisher nur möglich, im Feld »othertext« Hinweise zu Verwertungsrechten anzugeben. Dieses Feld ist allerdings ein Freitextfeld, das auch für sonstige Metadaten genutzt werden kann. Mit dem Text type Code 27 ist es aber möglich, Subrights (»Description for subsidiary rights department«) im Feld »othertext« zu codieren. Monografische Netzpublikationen sollten mit folgenden Produktnummern versehen werden: Für ONIX for Books ist die ISBN-13 dringend empfehlenswert, zusätzlich sollte auch ein DOI vergeben werden. Die in Vorbereitung befindliche ONIX Version 3.0 wird für digitale Produkte optimiert bzw. angepasst, eine Veröffentlichung der neuen Version ist jedoch noch nicht terminiert. Die deutsche und die internationale ONIX-Mailingliste wird intensiv genutzt, um Fragen zu diskutieren und um neue Elemente und Felder anzumelden. Dort wird auch über die aktuellen Entwicklungen des Standards informiert.

ONIX-
Titelmeldungen

Im Anschluss an den Vortrag von Martin Lünig zeigte Susanne Puls (DNB), wie die an die DNB gelieferten VLB-ONIX-Datensätze im Katalogisierungssystem der DNB umgesetzt und weiterverarbeitet werden. Alle an das VLB gemeldeten Titel werden automatisch von der MVB an die DNB weitergeleitet. Anhand von Beispielen wurde gezeigt, wie in der DNB eine Titelaufnahme nach dem derzeit gültigen Regelwerk für elektronische Publikationen (RAK-NBM) erstellt wird. Die von der MVB gelieferten Metadaten dienen als Grundlage für die Katalogisierung von Netzpublikationen, sie werden anhand der Vorlage überprüft und gegebenenfalls aktualisiert, damit die Auffindbarkeit und Identifizierung gewährleistet ist. Zusätzlich werden verschiedene Codes, die zum Teil interne Prozesse steuern oder der besseren Selektion der Titel dienen sollen, und fehlende Angaben in der Titelaufnahme ergänzt. Bei der Katalogisierung wird außerdem eine Verknüpfung des Titels mit der Personennamendatei (PND) hergestellt, sodass alle zu einer Person zugehörigen Titel gefunden werden können; ebenso können Verknüpfungen zur

Gemeinsamen Körperschaftsdatei (GKD) und zur Schlagwortnormdatei (SWD) hinzukommen.

Eine wichtige Angabe für die DNB ist die URL der Netzpublikation, die entweder direkt auf die Netzpublikation oder auf eine Frontpage verweist. Neben der URL werden in diesem Feld weitere Angaben zur Netzpublikation, wie z. B. Zugriffsart, Dateiformattyp, Dateigröße angegeben. Das Feld ist wiederholbar und somit ist es möglich, mehrere URL zu einer Netzpublikation anzugeben. In der Titelaufnahme wird außerdem ein Link auf das in der DNB langzeitarchivierte Exemplar auf dem Archivserver der DNB ausgewiesen. Sofern im ONIX-Datensatz kein URN vorhanden ist, wird ein URN von der DNB vergeben, da dieser für die DNB obligatorisch ist; er stellt das Bindeglied zwischen Archivierungssystem und Katalogisat dar.

Im neuen Gesetz über die Deutsche Nationalbibliothek wurde der Sammelauftrag auf unkörperliche Medienwerke, d. h. Netzpublikationen, ausgeweitet. Die Pflichtablieferungsverordnung (PflAV) wird in den nächsten Monaten die Pflichtstückverordnung ablösen und Einzelheiten zur Sammlung

URN als
Bindeglied

FAUST

Archiv
Medienarchiv
Museum
Dokumentation
Bibliothek
Dokumentenverwaltung



MEHRDIMENSIONALE DATENBANK • RETRIEVAL • DOKUMENTENMANAGEMENT

- individuelle Datenstruktur
- umfassende Recherche und Navigation
- Bild- und Medienarchivierung
- Rechtschreibprüfung, Schrifterkennung (OCR)
- Datenqualitätssicherung, freier Report
- Intranet, Internet,
- Import, Export, Downloading
- u. v. m.

Alle Infos im Netz:
www.land-software.de

Postfach 1126
90519 Oberasbach
Tel. 0911- 69 69 11
info@land-software.de



**LAND
SOFTWARE
ENTWICKLUNG**

Neue
Sammelrichtlinien

regeln. In den neuen Sammelrichtlinien werden die Einschränkungen der Sammlung weiter ausgeführt. Sie werden nach Inkrafttreten der Pflichtablieferungsverordnung im Laufe des Jahres 2008 umgesetzt. Die Sammelrichtlinien bieten dem Ablieferer Informationen darüber, was die DNB von ihm erwartet. Die zentralen Aussagen zu Netzpublikationen in den Sammelrichtlinien sind:

Die DNB sammelt

- im Prinzip alles, was in öffentlichen Netzen angeboten wird,
- sowohl Netzpublikationen in Analogie zu gedruckten Medien als auch webtypische Erscheinungsformen wie Websites, Datenbanken,
- keine Netzpublikationen, die nur geschäftlichen oder innerbetrieblichen Zwecken dienen oder nur privaten Nutzergruppen zugänglich sind.

In einer ersten Stufe ist die Sammlung aus praktischen Gründen auf monografische Netzpublikationen eingeschränkt. Die Sammlung von zeitschriftenartigen Publikationen, Websites und Datenbanken wird in einer darauf folgenden Stufe bearbeitet.

Digital Object
Identifizier

Aus Sicht der Buchindustrie Standardagentur (BISAG) erläuterte Manfred Gravelius das Konzept des Digital Object Identifier (DOI). Ein DOI ist ein Persistent Identifier für digitale Objekte und wird in ONIX »Product Identifier« genannt. Zurzeit müssen zwei ONIX-Formate verwendet werden, um Metadaten an das VLB zu melden bzw. einen DOI zu registrieren. Ziel ist die DOI-Registrierung über ONIX for Books bzw. ONIX 3.0. Das Problem bei der bisherigen Version ist die Beschränkung auf die Angabe des DOI lediglich als zusätzlichen Produkt-Identifizier. ONIX for DOI Names deckt alle relevanten bibliografischen Metadaten ab und könnte deshalb für die VLB/DNB-Meldung genutzt werden. Deshalb ist ein Standard für Netzpublikationen auf Basis von ONIX durch Integration von ONIX for DOI Names in ONIX 3.0 nötig. Vonseiten des VLB und der Verlage wird angemerkt, dass sich die Publikationswelt in einem starken Wandel befindet. Der allgemeine Trend

geht zur Lieferung von Inhalten auf der Articlebene, statt wie bisher, zur Lieferung ganzer Zeitschriftenausgaben. Auch wird es nach Ansicht der Workshopteilnehmer immer mehr nicht abgeschlossene Publikationen geben, also verschiedene Versionen einer Publikation, die nicht mehr als Zeitschrift bzw. Artikel behandelt werden können. Als Wunsch wurde darüber hinaus die Bildung einer Arbeitsgruppe angeregt, die sich mit der Vereinheitlichung der zahlreichen unterschiedlichen Verlagsnummern und -identifikationen für verschiedene Dienstleistungen und Kommunikationswege befassen soll. Dies würde bei den Verlagen und deren Kommunikationspartnern die Workflow-Prozesse deutlich vereinfachen.

Die DNB wird mittelfristig alle Arten von Netzpublikationen sammeln, unabhängig von ihrer Publikationsform. Bis die technischen Möglichkeiten zur Sammlung dynamischer Publikationen geschaffen sind, muss es jedoch eine Beschränkung auf die bisher möglichen Publikationsformen, nämlich monografische Publikationen, geben. Der Geschäftsgang für zeitschriftenartige Publikationen wird derzeit überarbeitet, damit eine gegenüber dem jetzigen Verfahren einfachere Ablieferung möglich wird.

Ein Ergebnis des Workshops ist ein weit reichender Konsens über die im Metadaten-Kernset vorgeschlagenen Elemente. An wenigen Stellen wurden die Angaben im Kernset von der DNB spezifiziert und die Terminologie überarbeitet. Das überarbeitete Kernset findet sich auf der Homepage der DNB¹. Damit ist das Metadaten-Kernset als Mindeststandard für Datenaustauschprozesse zwischen Verlagen, der MVB und der DNB akzeptiert.

Auf Anregung der Verlage und auf Wunsch der Vertreter der DNB und der MVB soll die Kooperation zwischen der DNB und der MVB auf ein Meldeverfahren für Netzpublikationen ausgeweitet werden, das es den Verlagen ermöglicht, über eine Meldung sowohl die Dienste der MVB als auch der DNB zu erreichen.

Zukünftige
Sammlung aller
Arten von Netz-
publikationen

Workshop-
ergebnis ist
ein Konsens

Anmerkungen

1 <http://www.d-nb.de/netzpub/info/pdf/metadaten_kernset_extern.pdf>