

Verlage und Buchhandel

UWE JANSSEN

ONIX – Metadaten in Verlagen und Buchhandel

ONIX (the acronym for Online Information Exchange) is the first standard for enriched metadata about books, serial publications and other products in the book trade to be accepted around the world. It was originally developed in the U.S.A. with the initial goal of providing internet booksellers with a uniform format for catalog data and additional marketing material. From the beginning, book suppliers, bibliographic agencies and publishers in the U.S.A. (and soon thereafter in the U.K.) participated in the development of ONIX, supporting it with both funding and staff. The maintenance and continuing development of this new standard was then placed in the hands of EDI_{EUR}, an international umbrella organization for book trade standards. The Association of German Booksellers and the European Booksellers Federation are among the founding member organizations of EDI_{EUR}, whose seat in London is provided by the British publishers' and booksellers' organization BIC (Book Industry Communication). EDI_{EUR} was established ten years ago by book trade, publishing and library organizations in order to develop EDI standards (the acronym for Electronic Data Interchange) for the communication between booksellers and publishers, on the one hand, and between libraries and their suppliers, on the other hand. It approved guidelines for a series of EANCOM messages, including orders, confirmation notices, invoices and sales offers. A draft guideline for PRICAT messages (Price and Sales Catalogue) was developed several years ago as a method of transferring catalog data, but has until now never been tested, much less been put to productive use. In contrast, EDI transaction messages are frequently used both in the library world and also in the book trade throughout Europe.

ONIX (das Akronym steht für Online Information Exchange) ist der erste weltweit akzeptierte Standard für angereicherte Metadaten über Bücher, Fortsetzungswerke und andere Produkte des Buchhandels. Er wurde ursprünglich in den USA entwickelt, zunächst mit dem Ziel, dem Internetbuchhandel in einem einheitlichen Format Katalogdaten und zusätzlich Marketingmaterial zur Verfügung stellen zu können. Von Anfang an waren Barsortimente, bibliographische Agenturen und Verlage in den USA und bald auch aus dem Vereinigten Königreich an der Entwicklung von ONIX beteiligt und haben diese finanziell sowie personell gefördert. Die Pflege und Weiterentwicklung dieses neuen Standards wurde dann in die Hände von EDI_{EUR} gelegt, der internationalen Dachorganisation für Standardisierung im Buchhandel, gegründet und gefördert von Verbänden aus Buchhandel, Verlagen und Bibliothekswesen, darunter dem Börsenverein des Deutschen Buchhandels und der European Booksellers Federation. Büro und Sekretariat von EDI_{EUR} werden in London von Book Industry Communication (BIC), einer Gemeinschaftsorganisation der britischen Verleger- und Buchhändlerverbände, gestellt. EDI_{EUR} wurde vor zehn Jahren gegründet, um EDI-Standards (EDI = electronic data interchange, elektronischer Datenaustausch) für die Kommunikation zwischen Buchhandel und Verlagen einerseits und Bibliotheken und ihren Lieferanten andererseits zu entwickeln. Dafür wurden Richtlinien für eine Reihe von EANCOM-Nachrichten verabschiedet, darunter für Bestellungen, Auftragsbestätigungen, Rechnungen und Angebote. Ein Richtlinienentwurf für die Nachricht PRICAT (Price and sales catalogue), die für die Übermittlung von Katalogdaten bestimmt ist, wurde zwar vor einigen Jahren entwickelt, aber bisher noch nirgendwo in der Praxis getestet oder gar produktiv eingesetzt. Hingegen sind die transaktionsbezogenen EDI-Nachrichten sowohl im Bibliothekswesen als auch im Buchhandel in Europa vielfach im Einsatz.

BIBLIOGRAPHISCHE INFORMATIONEN IN VERLAGEN

Bisher haben bibliographische Daten für zahlreiche Verlage keine sehr große Bedeutung gehabt. Zwar melden praktisch alle Verlage ihre Titel seit Jahren schon an das Verzeichnis lieferbarer Bücher (VLB), jedoch geschah das in den wenigsten Fällen durch elektronische Übertragung von Daten aus einem Verlagssystem. Wenige, vor allem wissenschaftliche, Verlage in Deutschland haben sehr gut gepflegte Datenbestände über ihre Titel. In anderen Verlagen findet man häufig zahlreiche Daten in unterschiedlichen Systemen. Da sind für die Auslieferung relevante Informationen in einem Verwaltungssystem oder bei der Verlagsauslieferung gespeichert, für Marketingzwecke existieren Daten in Publishing-Systemen, mit denen Vorschauen, Kataloge usw. produziert werden. In den letzten Jahren sind zahlreiche Verlage dazu übergegangen, spezielle Verlagssoftware einzusetzen, die von der Planung bis zur Werbung und Auslieferung, Versand von Rezensionsexemplaren usw. alle Aktivitäten eines Verlages unterstützt. Damit ist die Möglichkeit gegeben, bibliographische Daten einmal an der Quelle zu erfassen und an alle Partner in der Lieferkette sowie an Bibliotheken, bibliographische Agenturen und Kataloghersteller auf elektronischem Wege weiterzuleiten. Es wird sich zeigen, ob Verlage jetzt und in Zukunft in der Lage sind oder sein werden, diese Daten in befriedigender Qualität zu liefern.

Bisher hatte diese Fähigkeit (oder die Bereitschaft dazu) keine große Bedeutung für den Verkauf der Bücher. Verlage mussten gute Vorschauen und Kataloge produzieren, um ihre Produktion dem Sortiment nahe zu bringen, gute Pressearbeit leisten, um möglichst viele Rezensionen ihrer Bücher zu erreichen, Autoren ins Fernsehen zu bringen usw. Im Zeitalter des Internets jedoch reicht dieses alles nicht mehr aus. Denn in einer Internetbuchhandlung kann ein potenzieller Kunde kein Buch in die Hand nehmen, um darin zu blättern, den Klappentext oder ein paar Seiten zu lesen, oder sich im Inhaltsverzeichnis zu informieren, ob was er sucht, in dem Buch auch behandelt wird. Im Internet muss im Gegensatz zur stationären Buchhandlung sehr viel mehr Information über ein Buch angezeigt werden, als in den klassischen Katalogen des Buchhandels traditionell enthalten war. Dafür reichen reine bibliographische Daten nicht aus, selbst wenn sie von wünschenswerter Qualität sind. Um ein



Uwe Janssen

die Erfassung bibliographischer Daten hatte bisher für Verlage keinen hohen Stellenwert

Buch im Internet optimal anbieten zu können, müssen zahlreiche weitere Informationen hinzukommen, z. B. Inhaltsverzeichnisse, Autorenporträts und -fotos, Textproben oder -auszüge, Rezensionen bis hin zu Video- oder Audioclips.

BIBLIOGRAPHISCHE INFORMATIONEN FÜR DEN BUCHHANDEL

Für den deutschen Buchhandel sind die Kataloge der großen Barsortimente KNO KV (Koch, Neff & Oettinger, Stuttgart und Koehler & Volckmar, Köln) und Libri (Hamburg) die bedeutendsten Nachschlagewerke, hinzu kommt in größeren Buchhandlungen das VLB. Alle haben gemeinsam, dass sie hauptsächlich textorientiert sind. Sie enthalten bibliographische Daten, und – zumindest auf der KNO-KV-CD-ROM – seit einigen Jahren zusätzliche Textinformationen zu einer wachsenden Anzahl von Titeln. In Zukunft wird auch der stationäre Buchhandel das Internet für die Titelrecherche nutzen. Der Katalog von KNO KV wird im Laufe des Jahres 2003 für die Kunden der Barsortimente online zur Verfügung stehen. Das ermöglicht dem Buchhandel, seinen Kunden dieselbe Information über Bücher anbieten zu können, die auch im Internet, z. B. im KNO-KV-Katalog unter www.buchkatalog.de abrufbar ist. Im Gegensatz zum Internetbuchhandel kann sie hier aber um das Wissen der Buchhändlerin und ihre Erfahrung, vielleicht auch Kenntnis des Kunden und seiner Wünsche und Interessen ergänzt werden.

Die Barsortimente unterhalten für die Produktion ihrer Kataloge umfangreiche eigene Redaktionen, in denen die Daten erfasst und für den buchhändlerischen Bedarf erschlossen werden. Die Katalogerstellung basiert hauptsächlich auf der Autopsie der Bücher, anhand derer die von den Verlagen gemeldeten Daten kontrolliert und ggf. korrigiert werden, sowie die Verschlagwortung nach einem einheitlichen Thesaurus und die Zuordnung der Warengruppen nach der gemeinsamen Systematik von KNO KV und Libri erfolgen, soweit letztere nicht schon vom Verlag vorgenommen wurde. Das VLB unterhält ebenfalls eine Redaktion, war jedoch in der Vergangenheit für die Erfassung der Daten auf die Meldungen der Verlage angewiesen, da eine Autopsie wie in den Barsortimenten, die ja die Bücher am Lager haben, hier nicht möglich ist. Durch die jüngst begonnene Kooperation mit Der Deutschen Bibliothek erhält das VLB durch Autopsie gewonnene Daten nun von dieser.

Die Meldungen der Verlage an die Barsortimente erfolgen in jeder nur denkbaren Form, nur – in Ermangelung eines geeigneten Standards – nicht in einem elektronischen Format, das eine maschinelle Weiterbearbeitung ermöglicht. Das VLB hat schon bisher

ca. 50 % seiner Daten auf elektronischem Wege erhalten. Dazu war den Verlagen eine Software zur Verfügung gestellt worden, welche die manuelle Erfassung der Daten auf einem PC mit anschließender Übertragung ans VLB ermöglichte. Die Barsortimente KNO KV und Libri haben darüber hinaus vor einigen Jahren gemeinsame Richtlinien für die Übermittlung von zusätzlichen Textinformationen und Umschlagsabbildungen in Dateiform herausgegeben, die allerdings vollkommen getrennt von der Meldung der bibliographischen Informationen abläuft.

Im Bibliothekswesen gibt es seit vielen Jahren allgemein akzeptierte Standards für bibliographische Daten, die allerdings für Zwecke des Buchhandels auf Grund der unterschiedlichen Aufgabenstellung nicht gut geeignet sind. Es fehlen z. B. für den Handel wichtige Angaben, wie Preise für verschiedene Länder, Lieferbarkeit oder Vertriebsrechte. Allen Katalogen und ihren Redaktionen gemeinsam ist die Tatsache, dass die von Verlagen gelieferten Daten bislang in den wenigsten Fällen ohne weiteres zu gebrauchen sind, da in vielen Verlagen die Lieferung exakter bibliographischer Angaben als nicht entscheidend für den Verkauf der Bücher angesehen wurde. Die Ausnahmen, einige wenige wissenschaftliche Verlage, bestätigen diese Regel.

Wie oben bereits ausgeführt, gewinnt aber gerade die Bereitstellung von umfangreichen Metadaten für den Verkauf von Büchern nicht nur im Internet zunehmend an Bedeutung. Zugleich sind in den letzten Jahren durch die technische Entwicklung die gedruckten Ausgaben der Barsortimentskataloge eingestellt worden, wodurch deren Verkaufserlös für die Finanzierung der Redaktionen weggefallen ist. Die Nutzungsgebühren, die für Online-Kataloge oder solche auf CD-ROM zu erzielen sind, können das nicht ausgleichen. In den letzten Jahren ist die Nutzung der Barsortimentskataloge und des VLB, die traditionell nur in Buchhandel und Bibliotheken benutzt wurden, auf das Internet ausgedehnt worden, so dass Bücher weltweit unabhängig von Lagerbestand einer Buchhandlung und deren Ladenöffnungszeiten verkauft werden können. Beide großen deutschen Barsortimente haben ihre Kataloge ins Internet gestellt, außerdem basieren die meisten anderen Kataloge der deutschsprachigen Internetbuchhändler einschließlich Amazon auf den Daten von KNO KV oder Libri oder gar beider. Die Daten des VLB sind unter www.buchhandel.de heute ebenfalls im Internet nutzbar.

Auch Bibliotheken und ihre Nutzer brauchen heute mehr als die reinen bibliographischen Daten. In den USA wurde schon verlangt, dass ein OPAC genauso aussehen solle, wie die Webseiten von Amazon.com.

die Kataloge der großen Barsortimente und das VLB sind hauptsächlich textorientiert

Ausweitung der Nutzung von Barsortimentskatalogen und VLB

Wenngleich diese sicher den Ansprüchen wissenschaftlicher Bibliotheken nicht genügen, sind dennoch viele Details auch in Bibliotheken sinnvoll nutzbar, von Autorenporträts über Inhaltsverzeichnisse bis hin zu (Leser-)rezensionen und vielem anderen. Voraussetzung für derartige Erweiterungen ist in allen Fällen, dass die benötigten Daten von Verlagen geliefert, bei den Katalogherstellern gesammelt und den Nutzern in elektronischer Form zur Verfügung gestellt werden. Die technische Möglichkeit dazu ist mit ONIX erstmals gegeben.

ONIX TECHNISCH

Ursprünglich nur für Bücher entwickelt, können mit ONIX heute auch elektronische Medien wie e-Books dargestellt werden. Mittlerweile gibt es auch eine Version für Fortsetzungswerke einschl. Zeitschriften. ONIX ist eine Spezifikation des Inhalts von Nachrichten. Sie beschreibt Datenelemente und definiert Codelisten. Neben der Beschreibung bildet die DTD (Document type definition) das Kernstück des Regelwerkes. ONIX ist also eine Anwendung der vom World Wide Web Consortium (W3C) empfohlenen Familie der XML-Standards (Extensible markup language). XML hat in den letzten Jahren sehr schnell weltweit große Akzeptanz und Verbreitung gefunden. »Obwohl XML ursprünglich als HTML-Alternative gedacht war, d.h. als eine auf Internet-/Intranet-Publikationen zielende Auszeichnungssprache, wurde rasch deutlich, dass es – ähnlich mächtig wie SGML, doch handhabbarer – als eine universelle Sprache zum Austausch, zum Retrieval und zur Verwaltung semantisch qualifizierter Daten dienen kann ... Schwerpunkt des Einsatzes ist und bleibt bis auf weiteres der Online-Austausch von Informationen.«¹ Dennoch handelt es sich um Text, d.h. eine XML-Datei ist auch für Menschen lesbar. XML Software ist mittlerweile in vielerlei Varianten überall verfügbar und damit preisgünstig, viele Produkte sind im Internet kostenlos erhältlich. Das ermöglicht es auch kleineren Verlagen, sie zu nutzen.

Die XML Tags, die ein Datenelement kennzeichnen, bestehen aus englischen Wörtern oder Abkürzungen, Werte sind in Codelisten definiert. Der Verlag eines Buches wird beispielsweise so angegeben:

```
<PublisherName>Springer Verlag</PublisherName>
```

Neben diesen »sprechenden« Tags gibt es eine häufig genutzte Variante von ONIX, in der kurze, nicht »sprechende« Tags benutzt werden, um das zu übertragende Datenvolumen zu reduzieren. Dasselbe Beispiel mit kurzen Tags sieht so aus:

```
<a197>Springer Verlag</a197>
```

ONIX for Serials wurde speziell entwickelt, um Metadaten für alle Formen von Fortsetzungswerken dar-

stellen zu können. Sie enthalten u.a. die komplette Produktbeschreibung, alle bibliographischen Details sowie eine strukturierte Beschreibung des Werkes auf allen Ebenen von der Zeitschrift über die einzelne Ausgabe bis zum einzelnen Artikel, Textinformationen wie Herausgeber, A&I Services, Bilder, Audio- und Videodateien, Links zu Webseiten, Preise und Lieferbarkeit in unterschiedlichen Märkten, Subskriptionsangebote für elektronische (e-) und gedruckte (p-) Versionen, Promotionen, Informationen über Anzeigen und Vorlagen. ONIX for Serials umfasst den Serial title record, adaptiert von ONIX für Bücher mit Kataloginformationen, Beschreibung des Werkes, Werbung und Vertriebsinformation sowie komplexere Informationen über Preise, Pakete, Bezugsoptionen in einem zusätzlichen Serial product record. Der Serial item record beinhaltet Informationen über die einzelne Lieferung als strukturierte bibliographische Information. Das Konzept des »serial item« ermöglicht Flexibilität über die Einheit »Ausgabe« hinaus, v.a. für e-Journals. Es ermöglicht umfangreiche Beschreibungen eines Artikels, einer Version eines Artikels oder eines Updates. ONIX for Serials liegt bisher nur als Entwurf vor, der gegenwärtig in Pilotprojekten getestet und danach offiziell veröffentlicht werden soll.

WEITERENTWICKLUNG UND NUTZUNG

Die Weiterentwicklung des Standards wird von einem ONIX International Steering Committee gesteuert, in welchem bisher nationale Benutzergruppen aus den USA, Großbritannien, Frankreich, Korea, Lateinamerika und Deutschland vertreten sind. Die laufende Arbeit der Dokumentation, Entwicklung der DTD und technischen Abstimmung mit den nationalen Gruppen wird von Beratern durchgeführt. Die deutsche Benutzergruppe existiert seit November 2000 unabhängig von Verbänden wie dem Börsenverein o.ä. In ihr haben sich Vertreterinnen und Vertreter der Barsortimente, von Verlagen, Softwarehäusern, dem VLB, Der Deutschen Bibliothek und Buchhändler zusammengefunden. Aus der Arbeit dieser Gruppe resultieren zahlreiche Ergänzungen, die Eingang in Release 2.0 von ONIX gefunden haben. Dazu gehören u.a. mehr Contributor Roles, Contributor Sequence Nos., die Reihenhierarchien des VLB, alternative Namen bzw. Transkriptionen, die Warengruppensystematik der Barsortimente, die Schlagwort-Normdatei Der Deutschen Bibliothek, unterschiedliche Euro-Preise für verschiedene Länder, die Meldenummern der Barsortimente u.v.m.

International hat ONIX sehr schnell, vor allem in den USA und Großbritannien, inzwischen auch in Australien und Neuseeland, Verbreitung gefunden. Dort wird es von den bibliographischen Agenturen wie

Standard für Metadaten
über Produkte des
Buchhandels

deutsche
ONIX-Benutzergruppe

R.R. Bowker, Nielsen Book Data, von Internet- und anderen Buchhändlern wie Amazon.com, Barnes & Noble, J. Bennett, von Verlagen wie Cambridge University Press, Wiley, Mc Graw Hill, Random House und vielen anderen, und von Bibliotheken wie der Library of Congress, der British Library und anderen Nationalbibliotheken benutzt. Die British Library hat gemeinsam mit der Firma OCLC ein Mapping von ONIX nach UNIMARC entwickelt, die Library of Congress ein solches für MARC21.

Die Popularität von ONIX zeigt sich auch darin, dass erwogen wird, es für die ISTC-Registrierung (International Standard Textual Work Code) zu benutzen, einen in Entwicklung befindlichen Standard für die Registrierung von Werken unabhängig von ihrer Veröffentlichungsform. Auch die International DOI Foundation (DOI = Digital object identifier) prüft den Einsatz von ONIX als Metadaten-Standard für DOI-Registrierungs-Agenturen.

Gegenwärtig (Februar 2003) ist das Release 2.0 aktuell, das im August 2001 veröffentlicht wurde, Release 2.1 ist in Vorbereitung. Darin werden zahlreiche Änderungs- und Ergänzungsvorschläge der nationalen Benutzergruppen Berücksichtigung finden, die jetzt weitgehend aus praktischer Erfahrung stammen und nicht mehr überwiegend aus der Theorie.

In Deutschland steht die produktive Nutzung erst am Anfang. Das VLB nutzt ONIX für sein neues Redaktionssystem und die CIP-Meldung an Die Deutsche Bibliothek. Die Barsortimente KNO KV wollen im Jahre 2003 erreichen, dass möglichst viele Verlage ihre Titeldaten mit ONIX melden. Elektronische Titelmeldungen ermöglichen es, die gemeldeten Titel unverzüglich in den buchhändlerischen Bibliographien anzuzeigen und zugleich für das Publikum im Internet. Damit ist jeder Titel unmittelbar nach der erstmaligen elektronischen Meldung recherchier- und damit bestellbar. Kurz vor bzw. bei Erscheinen des Buches können die Daten aktualisiert und ergänzt werden, ebenso später während der gesamten Lebenszeit des Buches. Für Verlage ist dieses nicht in erster Linie eine technische Aufgabe, sondern ein wichtiger Bestandteil ihres Marketings und der Werbung. Für die technische Realisierung sind dann die EDV-Abteilungen der Verlage bzw. die Softwarehäuser zuständig. Vielfach wird es notwendig sein, zunächst die in verschiedenen Formen im Verlag vorhandenen Daten so zu organisieren, dass die Umsetzung in ONIX ermöglicht wird. Mittlerweile sind die ersten Softwareprodukte für diese Umsetzung auf dem Markt erhältlich.

Allerdings werden elektronische Meldungen der Verlage die Autopsie im Barsortiment nie ersetzen können, da die Einheitlichkeit und Qualität des Ka-

talogs auch in Zukunft nur durch zusätzliche redaktionelle Bearbeitung der Verlagsdaten sichergestellt werden kann. Niemandem nützt ein Katalog, der nicht durch einheitliche Verschlagwortung erschlossen ist oder in welchem Autorennamen je nach Verlag unterschiedlich geschrieben werden. Der Standardisierung der Datenaustauschformate zum Trotz behalten die Katalogredaktionen ihre zentrale Bedeutung für die Erschließung der Literatur, die der Barsortimente und des VLB für den Buchhandel genauso wie die der Nationalbibliographien für die Bibliotheken.

INFORMATIONEN ÜBER ONIX

Die gesamte Dokumentation und die DTD sind auf den Webseiten von EDItEUR im Internet unter <http://www.editeur.org/ONIX.html> abrufbar, allgemeine Informationen über Aufgaben, Ziele, Aktivitäten und Sponsoren von EDItEUR unter <http://www.editeur.org>. Ein Diskussionsforum zum Thema in englischer Sprache wurde im Internet unter http://groups.yahoo.com/group/ONIX_IMPLEMENT eingerichtet. Hier werden hauptsächlich Fragen der Implementierung und der Technik behandelt.

Kontaktadressen für die deutsche Benutzergruppe

Dr. Nicola Klupsch, Springer Verlag,
e-Mail klupsch@springer.de
Uwe Janssen, Koch, Neff & Oettinger,
e-Mail uwe.janssen@kno-bs.de

¹ Michel, Thomas: XML kompakt. Eine praktische Einführung. München: Hanser, 1999, S. 20

DER VERFASSER

Uwe Janssen ist Leiter des EDI Clearing Centre bei Koch, Neff & Oettinger & Co. GmbH, Schockenriedstr. 37, 70565 Stuttgart, uwe.janssen@kno-bs.de