

**Association for Library Collections & Technical Services
Cataloging and Classification Section**

COMMITTEE ON CATALOGING: DESCRIPTION AND ACCESS

Task Force on ONIX International

Final Report

The Task Force on ONIX International has been charged with but not limited to:

1. Evaluating the relationship between library metadata (AACR2, MARC 21) and the ONIX International standard for representing and communicating book industry product information in electronic form to determine how well ONIX International maps into AACR2 and MARC.
2. Identifying the issues surrounding the use of ONIX International metadata in AACR2 cataloging records. The Task Force shall refer to the four user tasks set forth in the IFLA Functional Requirements for Bibliographic Records and the core record requirements established by PCC in evaluating the ONIX International standard.
3. Assessing the consequences and impact of integrating records containing the ONIX International metadata into library databases, evaluating mechanisms for integration, and recommending appropriate measures for libraries.
4. Preparing rule revision proposals and discussion papers as needed.
5. Monitoring of projects and activities that use ONIX International.
6. Investigating the feasibility of informing others of library perspectives through a designated liaison to the developers of ONIX International.
7. Informing CC:DA about the development of ONIX International, including the preparation of a summary of the standard which shall include the following information:
 - some background or history and community served
 - description of metadata element set
 - sample records if possible
 - citations for more information, implementation projects, etc., including Web sites.

Charge 1: Evaluating the relationship between library metadata (AACR2, MARC 21) and the ONIX International standard for representing and communicating book industry product information in electronic form to determine how well ONIX International maps into AACR2 and MARC.

ONIX stands for **ON**line **I**nformation **eX**change. It is a metadata standard developed by the publishing community as a standard means to exchange information about “book” product information electronically to wholesalers, e-tail and retail booksellers, other publishers, and anyone else involved in the supply chain. The American Association of Publishers (AAP) developed ONIX during 1999 in conjunction with the major wholesalers, online retailers and book information services. Intending to provide publishers a means of sharing product and supplier information usable on the Internet. ONIX was designed as a solution to two problems:

- the lack of consistency and standards in data exchange formats in use by book wholesalers and retailers and the need for a universal, international format in which all publishers could exchange information;
- the need for richer book data online since there is no physical book for the potential buyers to pickup and peruse on the Internet.

Based on EPICS (EDItEUR Product Information Communication Standards), the release of ONIX 1.0 in January 2000 is the culmination of the combined experience of the Book Industry Study Group (BISG) in the US and the Book Industry Communication (BIC) in the UK. ONIX release 1.2 was issued in December 2000. The most recent version, ONIX 1.2.1, released June 1, 2001, is known as **ONIX International**.

ONIX defines both a list of data fields and how to send that data in an “ONIX message.” ONIX specifies over 200 data elements with standard definitions. Each element is defined to ensure consistent use. Some elements are required (such as ISBN, author, title) and some are optional (such as book reviews, cover image). ONIX messages are written in XML using the ONIX DTD (Document Type Definition).

The ONIX standard is currently limited to exchanging information about print books. Efforts are underway to broaden the standard to include other formats, such as videos, serials, and electronic books, and to include additional information, such as the concept of digital rights. It is expected to be an evolving standard.

Publishers such as Cambridge University Press and Houghton Mifflin are currently using ONIX to send data to booksellers. Amazon.com, and Barnes & Noble are among the booksellers equipped to receive ONIX data from publishers. It is expected that the list of booksellers and publishers who have implemented ONIX will grow rapidly within the next few years. The Library of Congress and OCLC are both prepared to receive ONIX data, and hope to use it in the future as seed data for producing MARC records. Mappings from ONIX to MARC21 and to UNIMARC are both available. (See Charge 7.2 for more technical information on ONIX)

Charge 2: Identifying the issues surrounding the use of ONIX International metadata in AACR2 cataloging records. The Task Force shall refer to the four user tasks set forth in the IFLA Functional Requirements for Bibliographic Records and the core record requirements established by PCC in evaluating the ONIX International standard.

In 1994, a Task Group appointed by the Cooperative Cataloging Council, now known as the Program for Cooperative Cataloging (PCC), defined a core record standard. The core record standard specifies a minimum set of data elements below which the PCC has agreed program records will not go, although the standard itself can be used by any library. The core standard was established within a MARC environment and was designed specifically to be used within the context of the Program for Cooperative Cataloging's national cataloging program. However, the minimum set of data elements could be adopted by any metadata standard, including ONIX.

All of the elements are present in ONIX that are needed in a MARC record to make a core (or full) record except for type of material and indicators. ONIX would therefore be suitable for a core record or for a bibliographic record, as far as the data elements themselves are concerned. The usefulness would currently be limited to printed books, however, and there might need to be a mechanism developed that could serve the same purpose as indicators to facilitate the display of the data elements in a catalog.

However, a hallmark of the PCC core record standard which sets it apart from most other metadata schemes, is that all of the non-descriptive data elements, the access points, are required to be under authority control. The ONIX metadata standard does not have any such requirement and, therefore, would be considered deficient for use in most library catalogs without additional work being done to integrate them with authority-controlled records and collocate the headings. It should be noted that the ONIX standard does include many data elements which ordinarily reside in authority records rather than bibliographic records, such as some of the elements for Contributors, such as professional position, affiliation, biographical note, etc.

In September 1997, Standing Committee of the IFLA Section on Cataloguing published its final report on the Functional Requirements for Bibliographic Records. The four user tasks set forth in that report are: find, identify, select, obtain. As the ONIX standard has all of the data elements that are present in the core (or full) standard for bibliographic records, it can be said that ONIX records could be considered as successful in meeting these four fundamental user needs as the traditional MARC record. The one area of deficiency, already noted, is the lack of authority control that could make any of the user tasks more difficult or time-consuming.

Charge 3: Assessing the consequences and impact of integrating records containing the ONIX International metadata into library databases, evaluating mechanisms for integration, and recommending appropriate measures for libraries.

The assessment is divided into three sections: administrative data fields, descriptive data fields and access data fields (names and subjects/genres) RILIN and OCLC already accept publisher records in their databases with the assumption that these records are useful as a basis of cataloging, that is to say, it is better than cataloging from scratch. Libraries are accepting mark records created by vendors of aggregator databases of electronic journals, e.g. Ebsco. However, if the cataloger must remember to update headings, strip out unwanted descriptive information whether they be recorded in subfields in variable fields or fix fields, then it may be more bother than it worth in some cases. There has been a preliminary published by Celia Burton, "ONIX for libraries," which reports on a survey of British librarians on the usefulness of the ONIX fields (<http://www.bic.org.uk/onixlibrep.doc>) for cataloging. In the survey, British librarians were asked to rank the various fields according to whether they were essential or important. Low percentages in both suggest that fields lack usefulness in terms of management or meeting the functional requirements of the bibliographic record. What the survey did not ask is whether the field should be excluded all together from the libraries databases.

Administrative Data Elements

ONIX has 30 administrative fields, which includes the message header data fields. The fields that pertain to the control and management of the ONIX database but would not be useful to the rest of the bibliographic universe could be filtered out during the export process or as part of the importation to one of the utilities. Examples would include the header fields containing the DTD type, message release version, contact information for the publisher. Currently these fields are not included in the mapping. (<http://lcweb.loc.gov/marc/onix2marc.html>) These fields are necessary for internal control for the publisher but have little or no value to the library community: supply-to-country, order-time, pack or carton quantity, tax rate, price effective from/until. Some of the sales promotion information may be interesting but may be misleading if not continually updated: copies sold, Book club adoption. If these fields were filtered out before being added to the utilities, the libraries would not miss them. In the Burton survey, The supplier and trade data (except for price amount), sales promotion information and message header data rated very low in importance. The LC web document suggests that the supplier and sales data which do not have equivalents in MARC, could be mapped to locally defined MARC elements. It may be judicious to map only the price information, and filter out all the other information.

Access: Names, Subjects

The most problematic aspect of using ONIX records is the difference in the form of the heading within an access fields. One of the functional requirements of a bibliographical record is that it is possible to locate all the works belonging to a given author or on a given subject or title. However, if the forms of names, titles and subjects do not collocate in a library's catalog, it makes this task more difficult.

Names

Main entry is essentially ignored in ONIX. All authors are tagged "Contributor" and listed sequentially with their accompanying information. The personal names are given in their natural word order and inverted form. Corporate and Conference names are given in a single string in natural word order. According to the Pearson mapping, the first personal name will be added to the 1xx or main entry field.¹ In each case a scenario is given on how to map an unstructured name, structured, keyname. When there are multiple names, all subsequent names are sent to the appropriate 7xx field. This would be true even if all names listed were editors of a collection of essays written by various authors. Conference and corporate names are mapped only to 7xx fields. Under the current rules, the cataloger would need to re-evaluate the nature of the contributors in order to determine any changes in the main entry.

According to the LC mapping and Pearson table, affiliation information in ONIX field could be migrated to the #u of the 1xx or 7xx fields. Although this is a possible solution, it was surprising to find that it was not suggested filtering this information out. Knowing the author's affiliation can be useful information in evaluating the validity of the product, in some cases. However, affiliations may change between publications. In the Burton survey this information was ranked very low. Rather than map this information, it should be consider filtering out this information prior to its load in a national utility. In the examples supplied to us, none of the personal names had birth or death dates associated with them.

In order to create consistency between the publisher and the library files for headings that have already been established, one possible solution would be to allow publishers to access freely the name authority files. Another possibility is to attempt to machine match the headings with the NAF file. Names that matched more the one heading (for example, only the dates distinguish the names) would not be matched. There would still be the possibility of matching the heading from the publishers with the wrong heading in the NAF. However, this assumes that cross-references for forms valid in other countries would be available since publishers are international and provide "products" to libraries internationally.

¹ Pearson, Bob. Footnote 9 found at http://www.editeur.org/ONIX_MARC_Mapping_External.doc , p. 5. See also the record builder at the MARC standards site: <http://lcweb.loc.gov/marc/marc2onix.html#100>

Author biographical and series information is valuable for patrons and librarians to evaluate the suitability of the title. Unfortunately, in ONIX if there are multiple authors, all the information about all the authors is repeated for each author (see example records). The current mapping converts the affiliation, professional position and titles after the name to subfields in the 7xx field. The biographical information goes to the 545 note field (biographical information about individuals, history note about institutions or events), conference descriptions to a 500 note. Nothing was mentioned in the mapping or conversion process of deleting duplicate information. Burton's survey ranked biographical data on the author as important but not essential. Series title and numbering was considered essential by 88% of the respondents. The most current version of ONIX is 2.0 and post-dates the original mapping. In the latest release of ONIX, they have created a structure to handle series and subseries data.

In September 2001, ONIX mounted a draft documents concerning ONIX for serials (<http://www.editeur.org/onixserials.html>). The cataloging community may want to study the usefulness of these records.

Subjects

The ONIX standard allows for numeric classification and topical headings. Although the EPICS data dictionary allows tagging specific schemes, including LCSH, Dewey, and LCC, the Wiley example does not include these fields. Publishers may opt to use the BIC subject list (<http://www.bic.org.uk/subcat.doc>), which is substantially shorter than LCSH (111 pages vs. 4 thick volumes). If LCSH has been used, libraries would want to retain them. If the BIC list or local publisher subjects, it is assumed these would map to a 69x or 653 field. Many libraries have set up their local import tables to remove 69x upon record importation.

Descriptive fields

As noted above, some information associated with the access fields are added to the note fields. Edition, language, extent of item, audience, publisher place name and date can be mapped into appropriate MARC21 fields. Although the weight of the book has not been mapped into MARC, the width and thickness as well as the height are mapped into \$c in the 300 field. Since this information is rarely used for mass-marketed titles, it is surprising that this information has been retained. It leaves the burden on the cataloging community to update the field. Codes will need to be converted, as in the case of audience, supporting text in order to map to and display in the appropriate field. The ONIX metadata permits links to image/audio/video files. The links have been mapped to the 856 field \$u and the various codes being translated into a phrase in the \$3 or \$z. Since most libraries do not retain the book jackets, this might be interesting to some of the patrons. Another valuable source of information is information about any prizes the title has won (586 field). This can help the library collection developer and the patron evaluate the appropriateness of the work. Normally, this information is reserved for the belles lettres.

Summary

Bob Pearson's record builder can convert ONIX records into MARC21 records. However, the success of using records based on the ONIX standard hinges on the mapping of the fields from ONIX to MARC21, the ability of the utilities to filter out administrative fields, selectively filter out descriptive information, e.g., the weight of a book, and to match access points with authorized headings in the National Authority file and convert them to the authorized form. It may be advisable to tag the records as "k" so that libraries can update the publisher's record in OCLC so that all libraries can share in the upgrading of the record.

Charge 4: Preparing rule revision proposals and discussion papers as needed.

At this time, the task force does not recommend any rule revisions or discussion papers regarding ONIX International. Due to the current and active involvement of the Library of Congress and OCLC with this metadata standard, and the development of metadata crosswalks between ONIX and MARC, the task force feels that no further action is required.

Charge 5: Monitoring of projects and activities that use ONIX International.

Since the task force was formed in January 2001, we have been monitoring projects and activities that use ONIX International. We have done this primarily by monitoring the ONIX_IMPLEMENT electronic discussion group and other relevant electronic discussion groups, and by reading the announcements, meeting notes, etc. published on the Web sites of the sponsoring organizations listed under Charge 7.4 in this report².

Projects, standards, etc.

The American Association of Publishers (AAP) in the U.S., the Publishers Association/Booksellers Association Supply Chain Project, Book Industry Communication (BIC) in the UK, and EDItEUR (internationally) are the collaborators behind ONIX International, the international standard for representing and communicating book core content. More information about these organizations can be found in the sources listed under Charge 7.4 in this report, or online at <http://www2.hawaii.edu/~chopey/ONIXTFCharge7.4.htm>

The International DOI Foundation (IDF) is also participating in ONIX efforts. "We are working on a DOI Genre for ebooks, developed in association with EDItEUR/ONIX," reported Norman Paskin, director of IDF and NISO director. "The primary goal of this

² Also available online at <http://www2.hawaii.edu/~chopey/ONIXTFCharge7.4.htm>

work is to define the kernel metadata for a DOI ebook genre,” said Paskin. “An important secondary goal is to define extensions to the EPICS Data Dictionary, and to the ONIX International standard for book trade product information. This would enable ONIX to handle e-books.³

The Open Ebook Forum is an association of software developers, publishers, authors, information specialists and others that creates and maintains standards for electronic books. According to Rebecca Guenther, the Forum has created working groups to deal with the issues brought up by the AAP standards. These include the OEB Identifiers Working Group and OEB Metadata Working Group. Some of the participants in these new groups (from AAP) were responsible for the work that went into the proposed AAP standards. The two groups are evaluating the recommendations in the AAP documents, particularly will be looking at the relationship between ONIX (which is recommended for use in the AAP proposed standard on metadata) and Dublin Core. Dublin Core is used in the Open Ebook Forum’s Publication Structure document, which is a standard for the structure of the content of electronic books and also includes some high level metadata expressed in DC. Thus, that relationship is being explored (and it is not yet certain whether the AAP recommendations for identifiers and metadata will be accepted by the Open Ebook Forum working groups).⁴

There is no direct relationship between ONIX and the Open eBook standard. ONIX is a product information exchange standard, aimed at providing descriptive information, including bibliographic and trade data, on a range of product types handled by what might be broadly termed the book and serials supply chains. Open eBook, by contrast, is a product content exchange standard, aimed at delivery of eBook content in a standard (XML-based) format together with some descriptive metadata (for cataloging purposes) based upon Dublin Core.

The next release (1.3) of ONIX will include extensions for the exchange of eBook product information, but these extensions are independent of eBook format. The main connection between ONIX and Open eBook is that both will be viewed as sources of information for cataloging purposes. However, the relationship between ONIX and Open eBook is very much the same as the relationship between ONIX and bibliographic data in the front matter of a printed book. ONIX will generally provide advance information, which may or may not turn out to correspond to Open eBook metadata obtained from an “eBook-in-hand” (depending upon the diligence of the publisher in issuing new and revised ONIX records).

³ Terry, Ana Arias. “Ebook Frenzy: An overview of issues, standards, and the industry.” *Information Standards Quarterly* (October 2000).

⁴ Comments from an e-mail posted to the DC-GENERAL electronic discussion list by Rebecca Guenther, Jan. 9, 2001.

ONIX users and suppliers

Current users of ONIX International include online booksellers, wholesale and catalog booksellers, aggregators, publishers, abstracting and indexing services, secondary and linking services, bibliographic information agencies, systems vendors, standards bodies, and libraries.

Some of the earliest adopters of the ONIX International metadata schema were the major online booksellers. Amazon, BN.com (Barnes & Noble), Books A'Million, Borders, Fatbrain, Indigo, and Reiter's Books all employ ONIX International metadata in their Web databases. ONIX metadata is supplied to online booksellers by publishers who wish to sell their books and other media through the online bookseller. Publishers can either create ONIX-encoded bibliographic records themselves, or contract a vendor to do this work for them. Amazon.com has been accepting ONIX-compliant data from publishers including Wiley, Harcourt Brace, CUP, and Princeton since September 2000.

Among the types of metadata accepted by Amazon are synopses, descriptions, tables of contents, author interviews, author biographies, jacket images (flap and backcover copy), interior images, extracts, reviews, and sound clips.

Publishers who are creating ONIX-encoded bibliographic records themselves include Adobe Systems (Glassbook), Cambridge University Press, Columbia University Press, Harcourt, HarperCollins, Houghton Mifflin, John Wiley & Sons, McGraw-Hill, New York University Press, Paladin Press, Pearson, Random House, Time-Warner Trade Publishing (including Little, Brown), Princeton University Press, and Yale University Press.

One of the larger vendors supplying and distributing ONIX-encoded bibliographic records for publishers who wish to either submit records to an online bookseller or publish their own catalogs online or in print is a company called NetRead (<http://www.netread.com/>), which uses a product called "JacketCaster" to perform these functions for publishers and distributors.

The JacketCaster Web site lists these publishers and distributors as customers of the company's services:

- ABC-CLIO Publishing
- Barron's Educational Books
- Cambridge University Press
- Coriolis
- Duke University Press
- HarperCollins Publishers
- Harvard University Press
- INSPEC
- O'Reilly and Associates, Inc.

Princeton University Press
Small Press Distribution
William Andrews Publishing

Book Data is a company that creates and supplies ONIX-encoded bibliographic records to online booksellers. As of November 2000, Book Data had in their databases over 785,000 Short or Long descriptions, 235,000 Tables of Contents, 250,000 Market rights statements, 135,000 Jacket images. A “significant majority” of these had been created manually by Book Data’s editorial staff.

A British company called Whitaker also supplies ONIX records. Whitaker receives XML feeds from publishers and passes these on to online booksellers. The company is also a member of the ONIX steering group. Among the early contributors of ONIX data to Whitaker have been publishers Cambridge University Press, Harper Collins, Scholastic, and Orion. Whitaker is working with other publishers to get them interested in supplying ONIX data, and to help them in getting set up to create the data.

MUZE, a provider of “new and innovative ways to build custom e-commerce solutions that merchandise music, books and videos” (<http://muze.com/>) supports the use of ONIX metadata and builds it into applications that it develops for its customers.

Cambridge University Press has developed an in-house online data delivery system called DataShop. The project started June 2000 with an in-progress Oracle database and one dedicated programmer. DataShop was developed during the same time that the centralized bibliographic database system already in place in Cambridge was being gradually migrating to Oracle. This provided an opportunity to shape the database and DataShop around ONIX and EPICS (the data dictionary that ONIX is based around) simultaneously.

The Cambridge ONIX output comes from DataShop, which lets users create ONIX files across the full CUP bibliographic database of c.50k titles for either the UK or US market with short tags or readable reference names. Files can be delivered once or daily, weekly or monthly depending on customer preference. Files can be delivered via email or FTP. Both full file refreshes or incremental updates are available.

The data in the Cambridge ONIX feed comes via DataShop directly from its Oracle database which is updated daily from CUP’s VISTA commercial systems in Cambridge, North America, and Australia. Marketing data and cover images are also fed into the Oracle system daily. In addition to using their own DataShop system, CUP also uses NetRead’s JacketCaster application. CUP’s data feed to NetRead is already formatted in ONIX, so NetRead’s function is simply to distribute the metadata for CUP.

CUP has initiated a pilot project to digitize their book production workflow. They will be creating ONIX-like metadata for their titles from the point the copy-editor begins work on the type-script. This will provide better bibliographic data earlier in the book’s life,

leading to more accurate advance orders, which will in turn mean fewer returns and fewer costly reorders.

As of October 2001, the Gale Group was developing ONIX formatted data for title feeds to its UK bibliographic agencies. Gale Group currently supplies its agencies with MS Excel formatted data but the company's longer term plans are to supply ONIX formatted data.

It is clear from many messages posted to ONIX_IMPLEMENT that the ONIX standard is well known among even the smallest book publishers. These small publishers are grappling with learning how to create ONIX metadata as it is increasingly required of them by the various distribution channels they use to sell their books.

Book distributors, wholesalers, and others using ONIX in their e-commerce applications include Baker & Taylor, Follett, Ingram, Login Brothers Book Company in the US (and their Ernesto Reichmann Distribuidora de Livros Ltd. in Brazil), and National Book Network. Ingram is working with Syndetics (www.syndetics.com) on a catalog enrichment program.⁵

Informata.com, an Internet portal produced by Baker and Taylor, is working on making book reviews and other catalog enrichment features available through library OPACs. The metadata content will reside and be managed on Informata's server, and access will be provided to the local ILS systems. Libraries won't need to process this content. A demo of this functionality can be seen at www.youseemore.com. Innovative Interfaces and TLC are ILS vendors who are reportedly working on incorporating this functionality.

BookSense.com, the e-commerce product for independent booksellers produced by the American Bookseller's Association, provides co-branded websites at participating retailers' URLs that utilize ONIX metadata and the ABA's title database of more than two million books.

Suppliers of bibliographic information and publishing data that are using ONIX include R. R. Bowker, and Saur, which uses or will use ONIX records in its *German Books in Print* database.

Bibliographic utilities

As of September 2001, OCLC had a program (software) written that would convert ONIX data into MARC, so that the bibliographic records could be matched against WorldCat. This program has not been installed, as there is not, of yet, any demand for it. As OCLC prepares to go to a new database structure (ORACLE), they are looking at accepting other types of data for input in addition to MARC. Dublin Core is at the top of

⁵ More information about Syndetics and catalog enrichment can be found at http://www.loc.gov/catdir/bibcontrol/calcano_paper.html

the list, but ONIX is on the list as well. It could move up in priority on the list in the future depending upon what relationships OCLC develops with publishers, booksellers, etc. who are using ONIX to transmit data.

National libraries

The most recent report available (October 12, 2001) from the Library of Congress Cataloging Directorate's Bibliographic Enrichment Advisement Team (BEAT) says that the team currently has three ONIX-related initiatives: implementing a Text Capture and Electronic Conversion-like project to utilize data from existing files within the scope of BEAT interests: e.g., TOC; exploring the possibility of using ONIX for "outgoing" records with NewBooks (q.v.) distribution, etc.; and participating in other initiatives dealing with ONIX applications relevant to the Cataloging Directorate.⁶ There also seems to be interest from other national libraries in ONIX as a format for electronic CIP submission (and return of enriched data to publishers).

At the Charleston Conference in November 2000⁷, Cindy Cunningham from Amazon.com spoke about catalog content, libraries, and ONIX. She said that Amazon manages 10 catalogs for its separate product lines. Everyone's perceptions about data and metadata are changing. The value of data is becoming widely appreciated. Publishers don't want to have to reformat data for various vendors, e-tailers, and aggregators. Amazon wants to compete on presentation, pricing, and interface, not on the quality of their metadata. What can libraries learn from what Amazon.com has done? Libraries want content dealing with licensed reviews, lit crit, author bios, TOCs, indices, covers, publisher and author content, excerpts, first chapters, e-versions. See the author's review for the Anarchist's Cookbook. Libraries also could generate content, such as customer reviews, in-house reviews (i.e., selectors and professors), synopses, annotations, related readings, and recommendations and ratings. Customer reviews are a pain to manage and filter, but they generate tremendous buy-in from the reading community. Content also could include video streaming, audio clips, internal images, three-dimensional, rotatable images. Content pitfalls include matching content to items, controlling contributed material, version/edition matching, paying for good data, finding relevant data, getting consistent data from providers, and storing and archiving data. ONIX can provide a format that scopes out the content needs. It provides a guide for publishers to remind them of all types of needed/wanted metadata. ONIX could provide a consistent way to receive data that could be automated in-house. ONIX is being designed for a marketplace environment, not for a cataloging environment. There has not been consistent input from the library community as the ONIX format has been developed.

⁶ <http://www.loc.gov/catdir/beat/> (viewed 11/11/01)

⁷ Notes available at <http://pallus.cic.uiuc.edu/ciclibraries/MtgNotes/Charleston00.htm>

Charge 6: Investigating the feasibility of informing others of library perspectives through a designated liaison to the developers of ONIX International.

Due to the continuing and active involvement by the Library of Congress and OCLC in current ONIX initiatives, this task force does not recommend a designated liaison from CC:DA or ALCTS; only continued monitoring of developments and versions on the official ONIX website.

Charge 7: Informing CC:DA about the development of ONIX International, including the preparation of a summary of the standard which shall include the following information:

Charge 7.1: background or history and community served

See Charge 1 for this information.

Charge 7.2: ONIX Metadata Element Set

General Description

ONIX is a rich metadata scheme that is comprised of 235 elements of information that fall into 24 categories. The set includes descriptive, administrative and structural metadata elements. The level of granularity of information is finer than that has been developed by MARC/AACR2. Although most of the descriptive metadata elements map to MARC, many of the administrative and structural elements in ONIX do not have an equivalence in MARC. Examples of these non-equivalent data elements include:

- Weight [of the book] (Descriptive)
- Supplier and Trade data (only 9 of 30 Administrative elements map)
- PriceComposite (Administrative)
- Outof PrintDate (Administrative)
- Downloadable copyright notice (Administrative)
- ONIX message elements (Structural)

ONIX standards allow two levels of implementation: Level 1, or ONIX-Lite, is designed to support publishers who have not developed a database for information management. It does not require reference to the XML DTD (release 1.2, 5/11/2000 p. 5). Level 2 contains all of the elements of Level 1 plus elements that provide for more in-depth description. In version 1.0 of ONIX (01/2001), 42 of the 148 elements were considered "crucial." With release 1.2 (11/2000) the total number of elements rose from 148 to 236, 82 of which comprise Level 1. Release 1.2.1 (06/01) contains only minor corrections and a few new values for existing elements, but no new elements. Unfortunately, what complicates the situation is that different countries have defined different subsets of

elements belonging to Level 1 (release 1.2, 24 /11/ 2000 p. 6). 45 of the elements are core to all countries (denoted by *); US publishers have added 4 elements (\$); UK and other European countries add 2 other elements (#)

Distribution of Categories of Metadata (number of elements in each category)

Categories	Level 1 # of elements	Level 2 # of elements
Reference number, type Source	5	8
Product numbers	3	8
Product form	2	11
Series	3	5
Set	4	6
Title	2	7
Authorship	5	16
Conference	0	6
Edition	3	3
Language	1	2
Extent and other content	2	7
Subject	2	23
Audience	4	12
Publisher	2	8
Publishing Dates	1	4
Territorial Rights	2	7
Dimensions	4	7
Descriptions and other supporting text	2	11
Links to image/audio/video files	3	14
Prizes	0	5
Related Products	4	13
Supplier and trade data	14	30
Sales promotion information	2	5
Message header data elements	12	18
Total	82	236

Crosswalks from ONIX to UNIMARC have been created, (<http://lcweb.loc.gov/marc/onix2marc.html> [ONIX 1.0 mapped] and see also Alan Danskin report at <http://www.bic.org.uk/reporton.doc> and Bob Pearson's mapping http://www.editeur.org/ONIX_MARC_Mapping_External.doc) that support publisher data to be added to national utility databases (OCLC/RLG/BL). Although the mapping is not one-to-one, the presence of publisher records provides a basis for creating and sharing bibliographic information using the MARC record structure based on publisher information. Many of the ONIX fields reflect the level of inventory control, rights, and publication or release information that the book industry requires that the library does not. Much of the information that does not currently map to MARC would not be missed

by library catalogers. In truth, some of the elements that are mapped to MARC are only possible if Level 3 Cataloging (AACR2) is being performed. Mainstream cataloging for publisher trade publications normally only uses Level 2. Hence, information about an author's affiliations, professional position (<b046> to 100/700 \$u) or contributor's role (\$e) would not normally appear in a library's bibliographic record. The end-user, however, would appreciate the links to the reviews, abstracts, and prizes won by the titles that are available in the publisher's records. Now that the use of the URL has expanded beyond the 856 field, these are links that libraries may want to consider retaining. The differences in granularity between the two metadata schema means that although fields can be mapped from ONIX into UNIMARC, they can not "reconverted."

Mapping the fields, however, is only one aspect of sharing these records. The form of names and the controlled vocabulary for subjects, audience, etc. differ from those used by the library community. The level of subject access for publishers and bookstores is far more general than the library access. Personal names may be inverted or in natural word order.

Charge 7.3: ONIX Record Examples

These records are created for business purposes and therefore the information contained within proprietary databases precludes outside access. However, there are a few examples that are available on the Web. Alan Danskin at the British Library has provided examples of records that were originally created in ONIX and converted in UNIMARC (<http://www.bic.org.uk/reporton.doc> appendix, pp. 4-9).

Guidelines for Online Information Exchange (ONIX) January 2000 1.0
<http://www.publishers.org/onix/onix.pdf> includes examples of ONIX records (pp. 53-54 of AAP pdf document)

There is an example of standard 1.1, level 2 in XHTML
(<http://www.editeur.org/samples/950-731-260-9A.xml>) (requires downloading to view)

Clifford Morgan at WileyUK has provide records for our study purposes. Wiley has adopted ONIX as their standard and continue to use the BIC set of subject codes. ONIX allows the implementor to use either BASIC or BIC codes

Conclusions

A devil's advocate may argue that it is often easier to start from scratch than to convert a publisher record into a standard UNIMARC record with AACR2 serving as the guide to formulate the content, NAF for names, and LCSH to provide the subject headings. Granularity as well as content may reduce the usability of these records. However, we may be gaining information about publications that would be useful for a patron to evaluate the title: increased information concerning the reception of the book, more information about the author, more detailed information about its currency.

Charge 7.4: Citations for more information, implementation projects, etc.

I. General and Background Information

ONIX International homepage (Developed and maintained by EDItEUR jointly with Book Industry Communication and the Book Industry Study Group)

<http://www.editeur.org/onix.html>

Includes:

- Link to Download of ONIX Release 1.2.1 (latest release as of 5/30/01) Guidelines and XML DTD
- Links to downloads of previous releases
- ONIX FAQs (frequently asked questions about ONIX)
- Link to subscription page for ONIX_IMPLEMENT (the e-mail listserv for ONIX implementers)

Powerpoint presentations on ONIX International

<http://www.bic.org.uk/onixsem.html>

Links to 7 presentations (Powerpoint, Word, or .pdf) presented at the seminar *ONIX International: How Better Product Information Sells More Books* (London, 14 November 2000):

- What is ONIX (David Martin)
- Amazon and ONIX (Mo Jacobs)
- Whitaker and ONIX (Michael Healy)
- BookData and ONIX (Peter Mathews)
- Cambridge University Press and ONIX
- Harper Collins and ONIX (Graham Bell)
- Libraries and ONIX (Alan Danskin)

II. Technical Information:

ONIX to MARC 21 Mapping (Network Development and MARC Standards Office, Library of Congress)

<http://lcweb.loc.gov/marc/onix2marc.html>

Includes mapping table in ONIX data element order, and record builder for creating MARC 21 records from ONIX data.

ONIX to UNIMARC Mapping (by Alan Danskin of The British Library)

<http://www.editeur.org/onixmarc.html>

EPICS Version 3: EDItEUR Product Information Communication Standards

<http://www.editeur.org/epics.html>

The data dictionary upon which ONIX International is based.

III. Implementations of ONIX International

ONIX International Implementers Information (The Book Industry Study Group, Inc.)

<http://www.bisg.org/onix.html>

Project details and contact information at organizations that are implementing ONIX International.

IV. Related Projects

AAP Open Ebook Publishing Standards Initiative (Association of American Publishers)

General information about the project:

<http://www.publishers.org/home/ebookstudyinfo.htm>

.pdf documents on Digital Rights Management, Metadata, and Numbering:

<http://www.publishers.org/home/ebookstudy.htm>

Book Industry Communication

<http://www.bic.org.uk>

BIC, set up and sponsored by The Publishers Association, The Booksellers Association, The Library Association and The British Library, develops and promotes standards for electronic commerce and communication in the book and serials industry.

V. Similar or Supporting Standards, Specifications, Initiatives, etc.

PRISM (Publishing Requirements for Industry Standard Metadata)

<http://www.prismstandard.org/news/2001/0401.asp>

Press release announcing the release of Version 1.0 of PRISM and links to PRISM home page and documentation.

BIC BASIC Standards for Product Information

<http://www.bic.org.uk/bbinfo.html>

Another subset of the EPICS data dictionary

Open eBook Publication Structure

<http://www.openebook.org/oebps/history.htm>

Includes the specification (Version 1.0) plus links to FAQ and sample records.

ANSI/NISO Z39.82-2001 (Published April 9, 2001)

Available for free download in .pdf format at: <http://www.niso.org/>

The ANSI/NISO standard on Title Pages for Conference Publications. This standard describes data elements that publishers, authors, and editors should use to create title pages or chief sources of information for conference publications in all subjects, languages, and formats and will be extremely helpful in assuring the communication of conference information to interested reader.

Task Force Members:

Bradford L. Eden, Chair

Ann Caldwell

Michael A. Chohey

Susan M. Hayes

Carol Hixson

Gertrude Koh

Cynthia Whitacre

Mary S. Woodley