Understanding Critical Elements of E-books: Standards for Formatting and Metadata


March 21, 2012

Speakers: Laura Dawson, Pat Payton, and Graham Bell

Metadata: Without You I'm Nothing

Metadata Quality and its Importance in E-Book Discovery

Laura Dawson
Communications Chief
Firebrand Technologies
Online: 61%
In Store: 37%

own it.
Now what..?

Metadata as a functional map (best practices)

Acquisition/P&L
- Assign ISBNs to tradable products
- Assign title/author/price

Editorial
- Assign BISAC categories
- Create initial product descriptions

Production/Design
- Create cover image
- Assign page counts
- Assign format types

Marketing
- Create market-specific product descriptions
- Aggregate advance reviews
- Confirm pub date

Metadata repository

...Profit!

Pat Payton (Patricia.Payton@bowker.com)
Senior Director, Publisher Relations and Content Development
Bowker
an affiliated business of ProQuest

Not All Items Need ISBNs

- Examples of current usage on digital works
- Rules to remember
- Third Party Assignments

ISTC
- Definition of identifier
- Process for obtaining
- Examples of usage on digital works
ISBN Must Be Valid

Unique ISBN for eBook

PDF

EPUB

Rules to Remember

content + print format = unique ISBN

content + file format + usage rights = unique ISBN

PDF for library use
perpetual access
multiple use
limited printing

EPUB
sold by retailer to single
consumer
Rules to Remember

Enhanced eBooks = unique ISBN

Rules to Remember

vendor <> unique ISBN (but is allowed)

pricing <> unique ISBN (but is allowed)
Retailer A sells at $15.95
Retailer B discounts to $9.95
Retailer C tests at $3.95 for 30 days, then $15.95

Third Party ISBN Assignments

Guidelines for Third Parties

- Link new ISBNs to publisher supplied ones
- Maintain publisher metadata for consumers
- Provide sales feedback using original ISBN
Remember This
content + file format + usage rights
= unique ISBN

International Standard Text Code (ISTC)

Democratic Identifier
Anyone can apply for ISTC
– Publishers
– Librarians
– Authors

Process of Assignment
User submits request to registration agency with metadata
Single international database searched
New ISTC assigned
Existing ISTC number returned
ISTC returned to user
ISTC returned to user
**Universal Identifier**

*Breaking Dawn*

ISTC A03-2010-0000000F-F

- E-Book A03-2010-0000000F-F
- Paper A03-2010-0000000F-F
- Library Binding A03-2010-0000000F-F
- Audio book - Special Format A03-2010-0000000F-F

**Relationship Identifier**

ISTC relationship codes
- Related Compilation
- Revision
- Adaptation
- Translated
- Abridged version
- Annotated version
PDF + EPUB = Same ISTC

Same Content = Same ISTC

Enhanced eBooks = ?

Same ISTC or derivative ISTC

Depending on text of work

Remember This

same textual content = single ISTC
Resources

ISBN
• Can be purchased through MyIdentifiers.com
• BISG Policy POL 1101 at www.bisg.org

ISTC
• ISTCs are available from Bowker
• BISG Identifiers working group

About EDItEUR

• not-for-profit membership organisation
• develops, supports and promotes metadata and identification standards for the book, e-book and serials supply chains
• based in London, but a global membership of publishers, distributors, wholesalers, subscription agents, retailers, libraries, system vendors, rights and trade associations
• also provides management services to International ISBN and ISTC Agencies

About me

• 20 years experience at the point where publishing and technology meet
• joined EDItEUR in mid-2010, primarily responsible for ONIX development
• formerly worked for HarperCollins UK
About ONIX

- ONIX for Books is the international XML-based standard for communicating book industry product information in electronic form

Typical use cases

- publisher needs to provide information about its catalogue of products to a distributor, wholesaler, retailer or other supply chain partner
  - includes both current and forthcoming products
  - may cover basic product information and a wide range of collateral material
  - scope extends over the full lifecycle for book, e-book and other products – ie includes post-publication updates to price and availability

Typical use cases

- specialist data aggregator or other organisation providing data services needs to collect metadata from many sources, and redistribute consolidated data about products to supply chain partners
  - this may include managing the ‘authority’ of data received from multiple data suppliers

Typical use cases

- distributor, wholesaler, retailer or library needs to collect metadata from its suppliers to populate its internal systems
  - may include both internal catalogue and consumer-facing systems such as an online store or OPAC
ONIX for Books is a standardised message specification, not a database.

**Roots of ONIX**

- 1997 BIC Basic
- 1998 `<indecs>` project
- 1998 W3C XML specification
- 1999 ‘Online Information Exchange’ initiative from AAP Digital Issues working party
- ONIX developed by EDItEUR, originally in collaboration with BISG (USA) and BIC (UK)
- 2000 ONIX v1.0 and 1.1 – retired
- 2001 ONIX v2.0 – retired

**Current status**

- 2004 ONIX v2.1 – most widely deployed
- 2009 ONIX v3.0 – growing in importance
- Widely used in North America, Western Europe, Japan, parts of Eastern Europe, with early implementations in China and Egypt
- Used by small and large organisations
- Included in many off-the-shelf IT systems, but also plausible for in-house developers

**ONIX 3.0 documentation**
### ONIX 3.0 data elements

<table>
<thead>
<tr>
<th>Category</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Message details</td>
<td>message details, identity and authority</td>
</tr>
<tr>
<td>Descriptive details</td>
<td>trade classification, product parts, collection titles, titles, contributors, conference, edition, language, extent, subject, audience</td>
</tr>
<tr>
<td>Product identifiers</td>
<td>record details, product identifiers, product form, special features, packaging, physical size, usage constraints</td>
</tr>
<tr>
<td>Collateral details</td>
<td>supporting text, cited material, supporting resources, prizes</td>
</tr>
<tr>
<td>Content detail</td>
<td>related material, related works, related products</td>
</tr>
<tr>
<td>Publishing details</td>
<td>markets, market publishing details, suppliers, discounts, prices and tax, reissue details</td>
</tr>
<tr>
<td>Supply details</td>
<td></td>
</tr>
</tbody>
</table>

**Showing 13–24 of 841,539 results**
Usage restrictions

- ONIX 3 only – <EpubTechnicalProtection> and <EpubUsageConstraint>
- limits on print, cut and paste, concurrent use, TTS, lending...
- can specify which DRM framework is used
- important for retailers to make constraints clear to purchasers
- ...and so publishers need to be clear to retailers

E-book accessibility

- EDItEUR worked with IDPF, DAISY Consortium and advocacy groups
- controlled vocabulary for accessibility
- publishers should incorporate testing for accessibility into their QA procedures, and feed the results into their metadata records
- Detailed, practical advice in O’Reilly book from Matt Garrish and Markus Gylling
ONIX and MARC

- different, and more or less complementary
- both contain bibliographic information
- but ONIX contains many data elements that are not within the scope of MARC
- and the importance attached to elements varies
- OCLC and LC have demonstrated value of using publisher’s ONIX data to seed or enrich MARC records
- mapping cannot be purely syntactic
Metadata in EPUB 3

```xml
<metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
    xmlns:dcterms="http://purl.org/dc/terms/"
    xmlns:bisacsh="http://purl.org/ontology/bisacsh"
    xmlns:ddc="http://purl.org/ontology/dddc"/>
<dc:identifier id="pub-id">urn:isbn:9780001234567</dc:identifier>
<dc:title>Roseanna</dc:title>
<dc:language>en</dc:language>
<meta property="dcterms:modified">2011-09-28T09:25:16Z</meta>
<link rel="onix-record" href="http://www.harpercollins.co.uk/onix/9780001234567.xml"/>
</metadata>
```