Book metadata and identification:
Bridging the divide from print to digital

Mark Bide
Executive Director EDItEUR

EDItEUR

About EDItEUR

- **Not-for-profit membership organisation**
  - Our role is to develop, maintain and promote the use of standards in the book and journal supply chains round the world

- **Based in London, global membership**
  - Publishers, distributors, wholesalers, subscription agents, booksellers, libraries, system vendors, rights management organisations and trade associations
  - 100 members in over 20 countries
    - US, Japan, China, UK and throughout Europe

- **Governing board of national, regional and international trade organisations to provide strategic direction**

- **Provide management services for ISO standards**
  - ISBN, ISTC, ISNI

EDItEUR

Book industry standards and the physical supply chain

Metadata standards are not new to the industry

▸ Managing a huge catalog of products: the ISBN

  ▸ Unambiguous identification of "things that are for sale"

▸ Managing a huge volume of transactions: EDI

  ▸ X12, EDIFACT, Tradacoms

  ▸ Exchange of commercial transactions

▸ Managing a huge volume of metadata: ONIX

  ▸ Exchange of rich descriptions

  ▸ An essential tool as commerce started to move from the physical bookstore to online

EDItEUR

What is the International Standard Book Number?

- ISO 2108 (1970; most recent revision 2005)

- 13 digit numeric string

- Includes some – but often misleading – affordance

- What does an ISBN identify? A book?

- A class of books – **a product**

Book industry standards and the physical supply chain

Metadata standards are not new

▶ Managing a huge catalog of products: the ISBN

  ▶ Unambiguous identification of "things that are for sale"

▶ Managing a huge volume of transactions: EDI

  ▶ X12, EDIFACT, Tradacoms

  ▶ Exchange of commercial transactions

▶ Managing a huge volume of metadata: ONIX

  ▶ Exchange of rich descriptions

  ▶ An essential tool as commerce started to move from the physical bookstore to online

What is ONIX for Books?

▸ XML communication format for sharing **book industry product information**

▸ Originated 1999 by the American Association of Publishers

▸ Current status: v2.1 widely implemented, v3.0 growing

▸ Implemented in many countries throughout the world – most recently Japan, China, Egypt, Turkey

▸ Allows the communication of information about publishers' products throughout the supply chain – to distributors, wholesalers, retailers and other partners

▸ In many markets, data is collected from many sources and redistributed in consolidated form to supply chain partners

▸ Used by small and large organisations, included in many off the shelf IT systems

▸

What do these standards have in common?

▸ Unashamedly, they are all about <u>commerce</u>

▸ Metadata and messaging standards are not **simply** about discovery – they are required for all aspects of commerce
  ▸ Helping people to find and buy things is a key driver of ONIX distribution…but there is lots more in an ONIX message

▸ Commerce is not constrained by borders or language
  ▸ Standards reflect that reality

EDI*t*EUR

ONIX and language

- Language of the standard and of the supporting documentation is English – although many national groups have their own translations

- No constraints on the use of character sets or reading direction

  - Active implementations in Japan, China, Korea, Russia, Egypt, Turkey, Bulgaria

- The codes are a language-independent notation – identifiers for <u>concepts</u>

- When an ONIX message crosses borders, the tokens continue to convey the same meaning

EDI*t*EUR

<ProductComposition>00</ProductComposition>

Downloadable ebook

<ProductForm>ED</ProductForm>

EPUB

<ProductFormDetail>E101</ProductFormDetail>

Fixed format

<ProductFormDetail>E201</ProductFormDetail>

No online components

<ProductFormDetail>E202</ProductFormDetail>

<ProductFormFeature>

<ProductFormFeatureType>06
</ProductFormFeatureType>

OS requirements

<ProductFormFeatureValue>03
</ProductFormFeatureValue>

<ProductFormFeatureDescription>iOS 5.1
</ProductFormFeatureDescription>

</ProductFormFeature>

Primarily text

<PrimaryContentType>10</PrimaryContentType>

<ProductContentType>06</Pro

With both audio and video components

<ProductContentType>13</ProductContentType>

3.0

What metadata does ONIX for books communicate?

Identity and authority

- Record details
- Product identifiers

Descriptive, including

- Product form
- Classifications
- Titles
- Contributors
- Edition
- Language
- Subject
- Audience

Collateral, including

- Marketing resources
- Supporting text

Publishing, including

- Imprint and publisher
- Publication date
- Territorial Rights

Related material, including

- Related works
- Related products

Supply, including

- Availability
- Suppliers
- Prices
- Discounts

From physical to digital – a mixed economy

▸ Metadata and identity are the "lifeblood of ecommerce"

▸ The core challenge is the increased complexity….

▸ ….of identification, of description, of transaction

▸ Metadata is as complex as the world it seeks to describe…

▸ … "simplification" of metadata = loss of information

EDI*t*EUR

Industry systems are not designed to deal with this complexity

- ISBN is a **product** identifier – but has been used as the primary key of many systems that have nothing to do with products
  - Definition of "a product" has become more difficult
  - Hardback, paperback….ebook?
- The potential number of products has become an order of magnitude greater
  - How do you collocate all these different products?
    - A work identifier (ISTC)?
    - A "release" identifier?
- How far do we have to manage <u>instance</u> identification?
  - The equivalent of RFID – already required for management of DRM

EDI*t*EUR

Managing the metadata explosion

- All metadata is essentially about identity
  - Particularly if it is to be unambiguously machine-processable
  - Essential for a commercial environment
- Public identification systems are not primarily technical but social – agreed upon norms and processes
  - Unambiguous rules for what is identified
  - Unambiguous rules of granularity – when are two things treated as being "the same thing" and when as different
- To be useful, public identification systems require publically accessible registries – so that others can know <u>what</u> is being identified
  - Books in Print – registries are not always "freely available"

EDI*t*EUR

The creation and management of authoritative metadata is never costless

‣ Common, authoritative metadata databases, if they are well run and maintained, will save costs for everyone…

‣ …but inaccurate, inconsistent and out-of-date metadata may be worse than no metadata at all

‣ Traditional systems for managing metadata and identity in publishing are no longer viable

  ‣ We don't deal simply in products

‣ Metadata itself is a <u>service</u> not a <u>good</u>

  ‣ It needs to be managed on an ongoing basis, not just manufactured once

‣ "Metadata should be free" is too simplistic

  ‣ There are costs associated with metadata creation and management that <u>someone </u>has to pay

Some questions I would like to hear answered

▸ **eBook identification**

  ▸ What are the classes of referents we need to identify?

▸ **eBook metadata: in-band or out-of band**

  ▸ What should be embedded and what associated by external reference?

▸ **Convergence between commercial and library practice**

  ▸ Can we share metadata more effectively?

▸

EDI*t*EUR

Book metadata and identification:
Bridging the divide from print to digital

mark@editeur.org

EDI*t*EUR