2010 standardisation challenges
The view from EDItEUR

Mark Bide, Executive Director, EDItEUR
LITA Standardisation SIG: Boston 16 January 2010

EDItEUR

About EDItEUR

- London-based global trade standards organization for books and serials supply chains
  - Established 1991
  - Not-for-profit membership organization with 90 members in 18 countries
- Supporting and promoting metadata and identification standards for a wide range of constituencies: publishers, wholesalers and subscription agents, retailers, libraries, system vendors
- Provide the secretariat for International ISBN Agency
- Standards development and management
  - EDI
  - RFID
  - ONIX family of metadata communications standards

EDItEUR

ONIX – origins

- Originally an initiative from the AAP
  - Response to the requirement for "more metadata" to support online retailing of books
  - "Metadata is all there is"
- ONIX   e
  - Developed by EDItEUR in close collaboration with BISG (in the US) and BIC (in the UK)
- Now "ONIX for Books" – the foundation on which all ONIX standards have been built

EDItEUR

ONIX family principles

- XML
  - Common approach to encoding, validation
- Designed for global application
  - Permissive, open structures
  - Able to cover a wide range of use cases and to be adaptable to local use without compromising the core structures
  - Encourage localised and appropriate profiling for specific applications
- Reuse of key structures and semantics within and between message families
  - Common composites
  - Shared code values
- Separate message structure from code values
  - Easy update of code lists while maintaining backwards compatibility
  - Only when absolutely necessary (new "major release" like ONIX for Books 3.0) is backwards compatibility lost

EDItEUR

The complete family

- ONIX for Books (Release 3.0 – April 2009)

- ONIX for Serials

- ONIX for Licensing Terms

- ONIX for identifier registration
  - The only part of the family I am not talking about today….

EDItEUR

---

ebook identification:
a key challenge for 2010

What is the challenge?

- Lack of consistency in identification of ebooks in the supply chain
  - Different publishers taking different approaches
  - Differences of view between UK and US
- Why not one ISBN per product?
  - "We only "publish" one generic format (eg .epub) and assign an ISBN to that"
  - "We are not responsible for formats provided by third party intermediaries"
  - "We don't care whether or not different product formats are listed in bibliographic databases."
  - "Our hardware-led channels do not require standard identifiers and customers will find our books through their preferred platform."
  - "Our system requires us to manually create and manage separate ONIX records for each ISBN we assign."

Why is this a concern for EDItEUR?

1. Real concerns about the future for the supply chain
   - Identification is the key item of metadata – the key to interoperability
   - The ISBN has served the supply chain very effectively for 40 years…
   - …and there are no obvious alternatives
   - Life without ISBN? What about Books in Print, for example?
2. We have recently launched ONIX Release 3.0
   - Enhanced management of ebook data
   - Structures assume ISBN for each product – cannot describe products without ISBN
   - Already causing a steady stream of queries
3. Same set of problems with EDI messages
   - Ordering without an ISBN
   - Sales reporting without an ISBN

Some solutions being proposed

- Publishers use a single ISBN for all ebook products, but distinguish between them in metadata
  - Applies ISBN at a completely different "level of abstraction" to current practice
  - The problem of "metadata bloat" just gets moved from one system to another
- [in addition] Allow ISBN to be applied by distributors as well as publishers
  - Substantial challenges in metadata "linkage"– who is responsible?
- Introduce a new identifier at this level of abstraction, but keep the ISBN for products where it is needed
  - Similar to the music industry's Global Release Identifier (GRid)
  - Unfamiliar and difficult to implement
- Use ISBN at both product and release levels – something like ISSN-L
  - Unfamiliar, difficult to implement, risk of confusion; seductively simple, probably wrong

Next steps

- International ISBN Agency continuing to work to find consensus
  - Still strongly supportive of "one ISBN per product"
  - However, still need to resolve the key question: "What distinguishes one product from another?"
    - Technical file format
    - Channel
    - License terms
- Need next to look at *technical* requirements
  - Which of the technical options creates least problem in implementation, both short and long term?
  - What should any new identifier look like?
- EDItEUR attempting to ensure that we respond to real market requirements, while not undermining good identification practice…
  - We may have to make things possible in the short term that are not ideal in the long term

Communicating rights & licenses: from theory to implementation

ONIX for Licensing Terms

- An overall framework – a way of communicating about:
  - Permissions: things you **may** do
  - Prohibitions: things you **may not** do
  - Conditions: things you **must** do
- It's all in the verbs – what we call *Usages*
  - The same constructs can be used for all types of agreements and licenses
- Where it has been used
  - ONIX-PL
  - ONIX for RROs
  - ONIX for ARROW [and perhaps ONIX for BRR]

ONIX for Publication Licenses

▸ Expresses licences agreed between publishers, hosting services, libraries, and corporate users

▸ Makes core "usage" terms machine interpretable

▸ Let's see what this might mean….

The user view of RELI – a JISC project

The user view of RELI

What is the challenge?

▸ The "chicken and egg" conundrum
  ▸ The requirement is real – but only libraries can create the demand
▸ Expressing licenses in XML is a considerable discipline for publishers and everyone else in the chain
  ▸ There is a steep learning curve for everyone
▸ Expressing licenses in XML does not overcome licensing disagreements
  ▸ Indeed, in the short term, the opposite may be true
▸ There are substantial challenges in identification
  ▸ Of resource, licenses and users
▸ A license registry can be useful to an institution in a number of ways, as well as providing permissions data for users
  ▸ Storing all licenses in one place for access by library staff
  ▸ Enabling comparisons of licenses

Next steps

▸ Continuing to work on promoting the implementation of the standard
  ▸ We need to create demand in the market place – essentially from libraries
▸ Some significant new projects are possible
  ▸ Always seeking partners interested in developing pilot projects
▸ Seeking strategic input….this is a strategic issue for everyone, not just for EDItEUR

ONIX for Serials?
what's in a name

What is ONIX for Serials?

- A "toolkit" of individual and composite elements and content definitions for the construction of message formats for a variety of applications
  - Messages are grouped as "families" – specialised messages based on the same schema
  - All messages share common structures and code list values for ease of implementation and interoperability
- The family:
  - SPS: Serials Price and Subscriptions
    - A catalogue message
  - SOH: Serials Online Holdings
    - A listing of electronic serials holdings details
  - SRN: Serials Release Notification
    - Notification of publication (online or physical)

EDI*t*EUR

Recent activity

- We have turned substantial attention to subscription products in the last year, and expect to do more in 2010
- Particularly driven by ICEDIS
  - Price catalogue
  - Claims
  - Long term – all ICEDIS messaging in ONIX

EDI*t*EUR

What is the challenge?

- ONIX for Serials was developed (in collaboration with NISO) specifically to support Serials…
- …but now there are a lot of other products which libraries get on subscription…
- …and ONIX for Serials does not have the constructs to support them

EDI*t*EUR

Next steps

- EDItEUR needs to ensure that it has proper input on requirements for the development of "ONIX for *Subscription Products*" [will have to rebrand!]
- Working with NISO to re-establish the joint committee
- Please contact me or Todd Carpenter if you are interested in participating
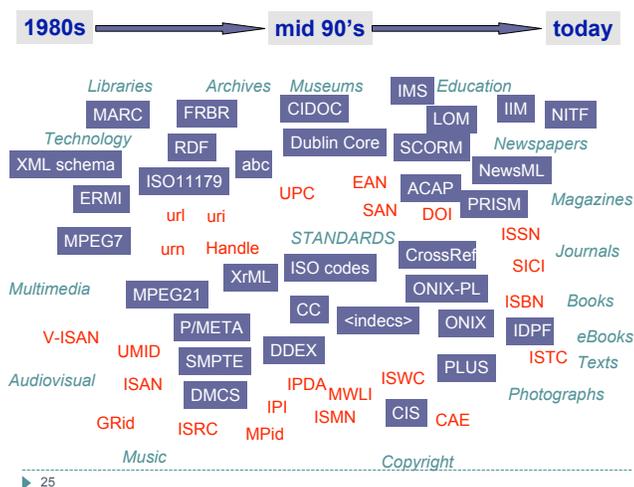 - tcarpenter@niso.org
  - mark@editeur.org

EDI*t*EUR

Convergence:
the next big challenge

What is the challenge?

"The nice thing about standards is that there are so many of them to choose from…"

Andrew S. Tanenbaum

EDI*t*EUR

Slide 1

Libraries   *Archives*   *Museums*   IMS   *Education*

MARC   FRBR   CIDOC   LOM   IIM   NITF

Technology   RDF   Dublin Core   SCORM   *Newspapers*

XML schema   ISO11179   abc   UPC   EAN   ACAP   NewsML

ERMI   url   uri   SAN   DOI   PRISM   *Magazines*

MPEG7   urn   Handle   *STANDARDS*   ISSN   *Journals*

Multimedia   XrML   ISO codes   CrossRef   SICI

MPEG21   CC   <indecs>   ONIX-PL   ISBN   *Books*

V-ISAN   P/META   ONIX   IDPF   *eBooks*

UMID   SMPTE   DDEX   PLUS   ISTC   *Texts*

Audiovisual   ISAN   DMCS   IPDA   ISWC   *Photographs*

GRid   ISRC   IPI   MWLI   CIS   CAE

MPid   ISMN

Music   *Copyright*

▶ 25

Slide 2 — So…what is the challenge?

- The process of proliferation has to stop…and be reversed
  - We need to seek more common ground
- This isn't easy…
  - Different requirements
  - Different history
  - Different understanding
  - Different legacy systems
- …but it is essential
  - Convergence is real
    - Between different applications
    - Between different media
- We cannot afford to maintain multiple solutions to essentially the same problem

EDI*t*EUR

Slide 3 — Next steps

- We need to find mechanisms for interoperability
  - Technical: check out the Vocabulary Mapping Framework project
    - http://cdlr.strath.ac.uk/VMF/
    - *The VMF matrix is a tool which can be used to automatically compute the "best fit" mappings between terms in controlled vocabularies in different metadata schemes and messages of interest to the educational, bibliographic and content publishing sectors.*
  - We need to make more progress on political and governance issues
    - A major focus for 2010

EDI*t*EUR

Slide 4 — Some conclusions

Slide 5 — Ten standards challenges

1. There is always a chicken and egg problem with standards implementation – and often little first mover advantage
2. Standards remove friction (cost) in the supply chain – but costs and benefits of implementation are not always "fairly" distributed
3. Dominant players can be either advantaged or disadvantaged by standards development – in the network world, this balance is changing; sometimes the rest of us will need to work hard to change the balance
4. While implementing standards will save you money, it will first cost you money – difficult at times of economic stringency, and difficult where an existing standard is "adequate"

EDI*t*EUR

Slide 6 — Ten standards challenges

5. Systems development resources are limited, and standards implementation is unexciting for development staff
6. Having great standards doesn't help if you don't manage your data properly – and good metadata costs money to create and maintain
7. There are some types of data – particularly rights and licence data – that as an industry we have no history of maintaining: weak system support, no clarity of responsibility

EDI*t*EUR

Ten standards challenges

Specifically for standards organisations

8. The timing of standards development is difficult – just in time can all too easily be just too late (or MUCH too early)

9. Standards development always suffers from "the tragedy of the commons" – it is always better if someone else funds it so you can use it

10. The standards community has to get its own house in order – we have too many industry standards organisations and media convergence means organisational convergence too

EDI*t*EUR

Thank you

mark@editeur.org

EDI*t*EUR