Data Exchange & Metadata Standards in the Journals Supply Chain

Tim Devenport
Lead consultant, EDItEUR

ALPSP Setting the Standard Seminar November 2015
Preview

• About EDItEUR
• Communications in the supply chain
• Transactional standards
• Descriptive metadata formats
• Are standards and good metadata optional?
• Best practice recommendations
• Emerging challenges
About EDItEUR

• Not-for-profit membership organisation
• Develops, supports and promotes metadata and identification standards for the book, e-book and serials supply chains
• Acknowledged centre of expertise on standards and metadata for the industry
• Based in London, but a global membership of 115+ publishers, distributors, wholesalers, subscription agents, retailers, libraries, system vendors, rights organizations and trade associations – located in more than 25 countries
About EDItEUR

• Also provides management services to International ISBN, ISTC, ISNI Agencies

• EDItEUR has five full-time staff, plus access to consultants from both the book and serials sectors

• We also work closely with other standards organisations, to ensure our standards meet the needs of their stakeholders too

• Member participation is vital to ensure that standards keep pace with evolving business requirements
Communication

• Clear, unambiguous, mutually comprehensible

• Automated and ‘machine-readable’ where appropriate – accuracy, speed, up-to-date

• And why?
  • End-user expectations
  • Accuracy
  • Immediacy
Types of standard

• Transactional standards
• Descriptive metadata formats
• Best practice recommendations
• … and mixtures of the above!
Transactional standards

EDItEUR develops or maintains three families:

- **ICEDIS** fixed-format standards, primarily used for agent <> publisher exchanges

- **EDIFACT** standards, used agent<>publisher but even more intensively agent<>library systems

- XML updates and extensions of the above
But standards for what?

• ICEDIS:
  • Two main flavours, first developed in the late 80s
  • AES (advice on existing subscriptions) synchronises records between parties
  • ORT conveys order, renewal or transfer instructions, also FTP addresses. Usage ubiquitous among journal publishers

• EDIFACT:
  • Some subscription orders, claims, invoices; also extensively used for library book supply
Transactional standards: ICEDIS & EDIFACT

Strengths and weaknesses:

• Uptake and established base for both is very extensive

• Work-horse formats for many years

• Largely developed pre-online products, packages, etc.

• Unstructured handling of person/organization details

• Legacy technological approaches: difficult to extend or amend
Transactional standards: Others

- Other standards or proprietary guidelines exist!
  - Including some old standards like X12 family in the United States
  - And a variety of PROPRIETARY formats introduced by a number of major players
- These clearly better than nothing, but they aren’t agreed or industry-wide, so players involved must cater for numerous formats
Descriptive formats

• Standard formats that describe objects, persons, organizations, agreements, license terms, ...

• As important as transactional standards, often precede them in supply chain choreography

• Some that EDItEUR looks after:
  • ONIX Books (including e-books)
  • ONIX-PC (product catalog for serials)
  • ONIX-PH (preservation holdings)
  • ONIX-PL (publication licenses)

• All EDItEUR descriptive formats are branded as ‘ONIX’ and are currently expressed in XML
Descriptive formats: ONIX-PC

• Seeking to communicate product and price detail for journals, journal packages & anything subscribable

• So, what does that include?
  • Title, Format, Publisher, Range, Price Models, Prices, Currencies for different Territories, Websites, for starters
  • And importantly, Identifiers for many of the entities above

• Early adopters include Springer NPG, Wiley Blackwell, Taylor & Francis. Systems vendor Publishing Technology recently announced it would make ONIX-PC an option available to its publishing clients
Other descriptive formats,

1: ONIX-PH

ONIX-PH (for preservation holdings):

- Created in conjunction with EDINA and its Keepers Registry project
- Keepers Registry tracks who has perserved what, in the e-journals sphere
- Partner archive organizations like British Library, e-Depot of the Dutch National Library, CLOCKSS, Portico, Hathi Trust and others
- ONIX-PH is a specialized holdings format – communicates what each archive has preserved, how much of it (ranges), to what depth, etc.
Other descriptive formats, 2: ONIX-PL

ONIX-PL (for publication licenses):

• A slightly special case
• NOT a license in itself, but a way of expressing licenses in machine-readable form
• So one could have e.g. an ONIX-PL expression of the CC BY Creative Commons license
• Extensively used here in the UK by Jisc
• But still awaiting wider adoption in US and beyond
the user introduce his own shorthand expressions.

**Metadata.** As important as being able to combine data elements to make composite data elements is the ability to associate explicitly with a data element a second data element which represents data "about" the first data element. This second data element we might term a "metadata element". Examples of such metadata elements are: an identifier, a domain "prescriptor" which specifies from what domain the values of the first element must be taken, an access code which limits the conditions under which the first data element can be accessed.

An alterable, prunable processor. How can we provide a wide variety of features, without: a) making the language processor too big and cumbersome, b) making the user learn a lot of information he doesn't need, c) unnecessarily restricting the user with lots of conventions about the use of the features which are provided. Simply providing a "kernel" lan-
‘In the future, adherence to standards will not be optional’
Best practice recommendations

• Essential complement to more formal standards
• Combination of accumulated wisdom, clear advice and elements of training
• Reflect views and requirements of multiple stakeholders in the process concerned
• Advice on business drivers, choreography, timescales and critically, WHY
• Sometimes, but not always, using standards or defined information items can be part of solution
Best practice in action: Transfer

- **The problem** – access interruption and delays when journals move between publishers
- Particularly but not exclusively, society titles
- **The response** – Project Transfer asked to investigate and propose improvements
- Started out as a UKSG Project, now maintained via a NISO Standing Committee
- Multi-stakeholder working group from the start
Best practice in action: Transfer diagnosis

- Fundamental problems threefold:
  - Poor or non-communication
  - No general agreement on key steps and timescales
  - Different perspectives of ‘Transferring’ and ‘Receiving’ publishers

- Potential hurdles – publisher worries about cooperation being construed as anti-competitive arrangements; conflicting contractual terms
Best practice in action: Transfer recommends ...

- A voluntary Code of Practice to which publishers are encouraged to sign up
- Now on its 3rd version and published as NISO RP-24-2015
- Declaration of a small group of information items by any endorsing publisher involved in a transfer
- Thus creating an underlying database ...
- ... on which a simple alerting service and blog are based
Best practice in action: What’s in the Transfer Code?

- Formalized roles & responsibilities for ‘Transferring’ and ‘Receiving’ publishers
- Timescales for specified actions and communications relative to an ‘Effective Transfer Date’
- How/when to transfer content, subscriber lists
- Guidance on identifier & URL handling
- Advice on existing licensing terms, preservation arrangements
**Transfer details**

**Receiving Publisher**
- Name: Taylor & Francis
- Contact name: Claire Washbrook
- Contact email: claire.washbrook@tandf.co.uk
- Transfer compliant: Yes

**Transferring Publisher**
- Name: World Council for Gifted and Talented Children
- Contact name: Thomas Tyler Clark
- Contact email: thomas.clark@wku.edu
- Transfer compliant: No

**Journal**
- Title: Gifted and Talented International
- Society: World Council for Gifted and Talented Children
- Print ISSN: 1533-2276
- Online ISSN: 0738-7849
- DOI:
- Frequency: 2

**Transfer details**
- URL: www.tandfonline.com/UGTI
- Effective transfer date: 2016-01-01
- First vol, issue, date: 31, 1, tbc

**Perpaccess policies**
- Receiving Publisher only will have archive
- Also, search the Keepers Registry (http://thekeepers.org/)
Another example: NISO’s ALI recommended practice

- Access and License Indicators
- (Originally ‘Open Access Metadata & Indicators”)
- NISO RP-22-2015
- Intended as an article-level mechanism to convey:
  - Whether or not the article is Open Access
  - Which license or usage terms might apply
  - Whether embargo periods are involved
The genesis of ALI

- Original mandate was to:
  - Specify metadata format and if possible a set of visual signals
  - Recommend mechanisms for distribution of this information
  - Assess feasibility of including statements of downstream rights
  - Support (or not) a specified set of use cases
ALI: eventual outcomes

• Eventual agreement on the concept of “free to read”, in order pragmatically to avoid endless philosophical/political debate

• Decision not to embed license information with article but rather to link or point out to it

• Two metadata elements specified, together with advice on how they should be populated and included in the article supply chain

• Visual indicators were eventually NOT specified, but they could in future be generated, based on the underlying metadata
ALI metadata elements

<free_to_read>

• indicates that content can be read or viewed at its current location by any user without payment or authentication

• has optional attributes start_date and end_date

<license_reference>

• to point to a public license or waiver, human and/or machine readable, that explains the terms for use or reuse of the content

• has optional attribute start_date
Distribution of ALI metadata

• Planned inclusion of the metadata elements in a number of existing metadata schemas and workflows, incl:
  
  - Journal Article Tag Suite (JATS), DOI registration metadata, ONIX, RDF, Dublin Core

• Strong encouragement for the license_reference to be populated with persistent URIs to ensure stability of link

• Side effect might be the gradual evolution of a ‘white list’ of frequently cited license URIs
A short tech excursion: formats & mechanisms

• Many existing and legacy standards have relied on transport of files, physically or electronically

• Some may persist – costs of change

• But current or future standards should be formulated to take advantage of:
  • Web services
  • JSON expressions
  • Public or secure APIs
  • And other emerging technologies
Emerging challenges: supporting Open Access

• Several pieces of work under way to better support Open Access workflows

• Jisc has very recently convened a group to look at this area and Sherpa/Romeo has been providing active support and advice for some years

• EDItEUR has undertaken an update and extension of the ONIX-PC standard to communicate journal-level OA features

• DOAJ is listing fully-OA journals, together with a series of relevant criteria or policy details
ONIX-PC extensions for Open Access journals

- Signal a change in the OA status of a journal
- Indicate which OA model (Gold, Green, etc.) is in use
- Describe whether the journal is fully or hybrid OA
- Define any embargo period(s) that may govern the OA status of articles in this journal
- Communicate details of the repository (or repositories) where OA articles for this are deposited
ONIX-PC extensions for Open Access journals, 2

- State which default license types are offered by the publisher for OA articles published in this journal and where license details may be found.
- Convey details of any article processing charges that may be applicable, including associated prices.
- Report the proportion of OA articles published in a hybrid journal during some defined reference period.
- List any directories or other authority sources in which a component is cited as being OA.
Some useful resources

- ICEDIS legacy standards: [http://www.editeur.org/131/Overview/](http://www.editeur.org/131/Overview/)
- Transfer: [http://www.niso.org/workrooms/transfer/](http://www.niso.org/workrooms/transfer/)
- ALI: [http://www.niso.org/workrooms/ali/](http://www.niso.org/workrooms/ali/)
- Contact me: tim@editeur.org