ONIX International Steering Committee Minutes

Wednesday 14th October 2015, 1:00–3:00 CEST (UTC+2),
Room Facette, Hall 3.Via, Frankfurt Messe

Attendees:

Luc Audrain (Hachette Livre)
Detlef Bauer (Libri)
Graham Bell (EDItEUR)
Emiel van Bockel (CB)
Maria Börman (Bokinfo)
Francis Cave (EDItEUR)
Michel Cervellin (BTLF)
Piera Constantini (Informazioni Editoriali)
Laurent Dervieu (ELECTRE)
Emad Eldeen Elakehal (ibiidi.com)
Fride Fosseng (Bokbasen)
Alain Fournier (BTLF)
Bente Franck-Sætervoll (Bokbasen)
Noah Genner (BookNet Canada)
Michael Gordon-Smith (Australian PA)
Alex Ingram (EDItEUR)
Uwe Janssen (KNV)
Tang Jiajun (TC505)
Chung-Su Kim (Korea Book Trade Promotion)
Anna Lionetti (AIE - MEDRA)
Karina Luke (BIC)
Giulia Marangoni (AIE)
Merve Okçuoğlu (Turkish Publishers Association)
Haeng Ung Park (Korea Book Trade Promotion)
Tuula Pelkonen-Tiri (Kirjavalitys)
Jesús Peraita (FGEE/DILVE)
Lisbeth Hankansson Petré (Bokinfo)
Simonetta Pillon (Informazioni Editoriali)
Marie Bilde Rasmussen (for Danish PA)
Darren Ryan (Deanta)
Howard Willows (Nielsen Book)
Wang Xia (RDG)
An Xiumin (TC505)

1. Welcome and introductions

Laurent Dervieu welcomed everyone to Frankfurt, and asked the participants each to introduce themselves. Graham Bell noted that we had a well attended meeting with nearly 20 countries represented.

2. Minutes of ISC Meeting held 15th April, London Book Fair, and matters arising

LD asked if there were any comments on the minutes of the previous meeting. GB noted that the minutes, as draft, have been on the website for a few months already. Howard Willows had a minor correction for page 7 where the minutes should have stated LV (for Len Vlahos) rather than LC. LD then called for and received approval of the minutes.

3. Report of current ONIX development work

GB circulated his activity report that covered the period from London Book Fair to present before the meeting. Just before London, the sunset date passed and support for ONIX 2.1 reduced. Schemas and tools were archived on the EDItEUR website, there was a change in tone of the response to queries by EDItEUR, and there have been updates to the training. All serve to emphasise ONIX 3.0 over ONIX 2.1, and the 'legacy' status of 2.1.

Levels of data requests (*ie* calls to the XML DTD and XSD files) have mostly declined since sunset occurred. Prior to sunset, there were 1m requests per month – roughly one every 3-4 seconds. These attempts have declined as people have realised that ONIX 2.1 has been archived and ceases to function. However, two bold entries show when the trend has not been followed, perhaps as someone new implements ONIX 2.1 without realising the DTD and XSD files must be installed locally.

At the time of the last meeting GB expressed concern at the levels of requests for the legacy XSDs which were misconfigured. Those also have reduced in recent months.

Sunset has caused fewer problems in the trade than we were expecting.

At this meeting last year GB noted ISC had agreed to a year of 'twilight' support for 2.1 codelist updates. This has continued through 2015, and has meant two or three dozen new codes in the ONIX 2.1 lists than would not otherwise have been added. 'Twilight' support ceases at the end of this calendar year. The codelists shown today will (almost certainly) be the final set of codelists to update codes for lists that are unique to 2.1.

Revisions in recent code lists have included further contributor roles to support comics, further educational support in classification and grade coding, clearer pricing of library or corporate content, pricing ability especially for 'subscriptions', and soundtrack details. In subscriptions, more work probably needs to be done, but many 'subscription' business models appear not to be commercially successful.

Alongside recent codelist updates, a slightly stricter XSD validation schema has been distributed for testing and comments. This has included more detailed checking of numerical elements, specifying numerical fields that may or may not have a zero, where real or only integer numbers are acceptable, and where only 0…100 ranges are valid. Given the feedback received (limited, but positive) these stricter definitions of numeric data elements will be included in the next release of ONIX 3.0. GB encouraged further use of the 'strict' XSD.

ONIX Acknowledgement message – GB has not heard of any further implementations as yet beyond the successful pilot between Hachette and the BNF (National Library of France), but GB reported he has had contact from people *intending* to implement in the coming twelve months. Luc Audrain confirmed the pilot has continued with the final release version of the Acknowledgement.

ONIX Training – the ONIX training programme has continued in the UK and US. The planned training programme in France has been delayed but is still being progressed.

Training material has been developed for online use by Alex Ingram. This has been expensive in time to produce, and therefore cannot immediately be free to use. It has been licensed to a UK

association, and it cannot be further licensed inside the UK until Frankfurt 2016, but outside the UK there are several interested potential licensees.

Schema.org – GB described this as a type of SEO, improving the effectiveness of data for search engine usage. However, it is important that publishers, retailers and others should not have to maintain this data *separately* from ONIX (thus increasing costs), but that the same data should be used for *both*. A BISG working group chaired by GB is documenting and will promote a mapping from existing ONIX data to Schema.org data – enabling the re-use of existing data to improve search engine rankings *etc*. Work is about half-way through, with documentation phase underway. The intention is to publish a joint BISG / EDItEUR recommendation.

LA asked if the schema.org work would cover not only the vocabulary but also the tagging. GB confirmed the documentation would cover examples of how the schema.org markup should be used within a real-world web page, and would where necessary either suggest extensions to the schema.org vocabularies or use external vocabularies to extend the expressiveness of schema.org to cover books. GB and BISG are keen to develop a quick and straightforward publication as there is a clear need for usable documentation.

Following initial discussion at the previous meeting of the ISC, work has begun on ONIX 3.0.3. GB has been having informal discussions about 3.0.3, and has collected a wide range of proposals from many parties. A separate discussion paper will be considered later in the agenda, but GB advised that the need was to balance the importance and value of the additions to ONIX 3.0.3 with the complexity and cost of supporting those additions, especially for *recipients* of ONIX.

All the proposals are for optional additions to the standard, to maintain compatibility.

GB noted that potentially every proposal may well have a keen supporter present. And while discussions can continue well beyond this meeting, some key decisions need to me made by the ISC now.

Translations – GB noted that EDItEUR has been approached recently by a company in Brazil which had translated a section of the ONIX Specification into Portuguese. Ideally, EDItEUR would prefer that people ask *before* commencing translations, but it always welcomes such activity. GB was able to spot some changes to make within the translation. This (partial) Brazilian Portuguese translation has now been posted on the EDItEUR website.

GB noted the French translation hadn't been updated since 2010, and thus omitted a few corrections made to the Specification in late 2010 and the additions made in 3.0.1 and 3.0.2. LD noted the concern.

The Best Practices Guide continues to be updated, usually alongside new codelist releases.

GB had fielded a query from the NL national group as to just how 'global' the best practices document really is, and what the remaining role for 'local' or national best practices is at present. GB noted he had described the global guidelines as being a base line, from which improvements can always be made – and that they are recommendations, not a set of rules.

There are occasionally variations in trade operations, and so national best practice still have an important role, but they should build on the global guide wherever possible. Important variations might well be inserted into the global guidelines as a note or sidebar if national groups inform EDItEUR.

Emiel van Bockel noted that he saw a lot of demand for an 'ONIX for dummies' due to the number of new people coming into the trade to work on tools without book trade experience. GB noted that there is a great deal of introductory information available, but the online training could be very helpful for this. EB noted the issues arose in particular with retailers, who are concerned at finding easier ways to get started. Often the retailers are then using outsourced resources in markets like India, and there are problems in getting the documentation to support that. GB acknowledged the problem, noting the many IT systems in the market that help publishers create ONIX, but that few help ingest ONIX into a retail system. LA noted that woocommerce – a popular e-commerce plugin for WordPress – supports the ingestion of ONIX 3.0. GB mentioned Firsty Group provide a similar toolkit, but agreed there was a relative lack of systems that are easy to pick up and use out of the box.

LD further commented that perhaps something branded and explained as a 'for dummies' approach to ONIX could have global application. He noted that he also saw the issue particularly for retailers in finding material pitched at the right level for their developers to start with. GB suggested that perhaps it could be that the online training which had been pitched at small publishers could be reoriented with a modest amount of effort to introduce ONIX concepts gently and simply to a developer audience. AI noted that the online training material actually has a dummy store to show how ONIX data might be represented by a retailer, but agreed with GB that the concern would be making that work both economically and as part of wider documentation for EDItEUR.

LA asked about the possibility of a reference implementation of ONIX ingestion. This is happening in the EPUB world, where Readium is (in effect) a reference implementation of an EPUB reader[1]. It does not provide a complete product, and developers still need to provide their own user interface among other elements to create a functional reader. Having something similar would clearly be a great assistance to users of ONIX, but it should be limited – perhaps by showing a method for creation of ONIX by XSL transformation from a flat file for particular elements such as bibliographic data or prices. GB replied that he would love to be able to do that, but it is a question of resources and the scale of the project. EDItEUR would have to look very carefully at how such a large project could be implemented.

Any reference implementation would have the specific issue of which technology to target (which DBMS, which programming framework *etc*). EB suggested that things be focussed upon PHP and MySQL, but others noted alternative frameworks would be equally valid choices.

[1] In fact, Readium is in part a reference implementation, but is also a commercially-licensable SDK that produces a very significant income stream to support ongoing development (see http://readium.org)

LA suggested this was a chicken and egg problem: no reference implementation, so we write more guidelines and best practice to improve consistency. New parties could be nudged in particular ways. A reference implementation should focus on basic parts of the record – title, contributor, cover, series, territory, prices, markets. Not to solve everything but to help people to build upon and learn from rather than a complete off-the-shelf product or SDK..

FC noted that such projects have been started before on GitHub *etc*, but none have reached a critical mass. Perhaps this should be a collaborative project? It is also worth remembering that something like price is 'basic' in the sense of 'fundamental', but not 'simple' – any full reference implementation would be hugely complex.

EB suggested that the problem is that the skillset of people implementing ONIX is often wrong.

> [In another context, GB has written "ONIX is regarded as complex – though most publishers and retailers use only a part of the full capabilities … Meanwhile, the problem is perhaps not the syntactic simplicity or complexity of the data format … , but the challenge of ensuring that those designing systems understand the semantic complexity and the complexity of the business itself, and that the user interfaces … help the data creators and the data users (the retailers and readers) to do so."]

LA also asked about mapping from ONIX to a 'flat' file structure. JP noted that unless the ONIX is produced from flat data, then there is no easy way to map ONIX to a flat file. (And several proposed ONIX 3.0.3 additions would make this even more complex.) Composites in the ONIX XML exist exactly because the underlying data required by business partners is far more complex than a flat file option could support.

GB asked whether there are two things here? A reference software implementation, and the definitions (*eg* 'what is an imprint?') that are part of the shared understanding within the industry. The latter details are within the documentation, but EB noted that people won't *find* the details. It's not easy enough. The need is to make the first step as easy as possible. There are real complaints in NL about this, and it is a significant barrier to adoption. GB plans a meeting with EB on this topic, but also keen to see further discussion more widely on reference implementations and documentation.

> [NB there is a glossary within the *ONIX 3.0 Implementation and Best Practice Guide*.]

EE confirmed that for anyone to implement ONIX *in full* is a huge job. Nobody does that. Encouraging people to go via flat files does not help with ONIX – the need is to move people up from flat file to a relational model.

BFS related her experience in Norway – this is about providing lots of detailed information and training. It's all about training. Offer to take full-day meetings, sit with them. Show them the important elements. It does vary from market to market, and on your exact business requirements. There are always some ONIX fields more important than others.

EB stated he already has one person full-time doing this, says it's a different thing (BFS disagrees), and CB cannot afford this long term.

SP noted that to keep the customer, she has to make flat files in some cases, for those customers who cannot take ONIX. These are often unique, and can be expensive (in time) to produce.

GB agreed, and pointed out that the issue often faced with flat files is that people want data that is simpler than the real world actually requires. We already know things are complex and then there are people in the business who want more detailed pricing and markets and so on. It's not getting simpler, it's getting more complex. Flat files are becoming more and more unrealistic. This is why a reference implementation is tempting. GB agreed on the need to think about this, and requested opinions via email.

4. Proposal for ONIX for Books Codelists Issue 31

GB introduced this by noting that there are two methods by which the standard can be enhanced and extended, and this plus the next agenda item cover each of them.

First, this proposal is about adding entries to codelists, which are revised on a regular three-month schedule. Codelist proposals are for additions and some minor refinements or modifications within the existing XML data structure. The next agenda item proposes additional XML tags and structures for the standard.

GB then summarised the codelists additions and refinements as proposed for Issue 31, which had been circulated to the national groups prior to the meeting, and called for comments.

EE – queried how many potential codes are available in 139. The codes are three alphabetical letters so 26^3 combinations (more than 17,000, or about 14,000 ignoring I and O).

The List 165 additions are to support features of the new VLB-Tix system, and there may well need to add others to list 165 in future.

FF asked whether it would be possible to add podcasts and instructional videos to List 158. GB noted that he had this proposal to hand (it will be included in proposals for issue 32).

GB noted some confusion over whether the French eco-participation tax was one tax or two (one aspect of the tax covers electronic devices, and the other covers 'furniture' such as display stands). A single code has been proposed, and it may need to be supplemented later.

HW asked about List 17. The original UK request was for a role 'Curated By' to describe a contributor who selected photographs for a book, but that has been worded as 'Non-text material selected by'. GB explained that in the UK 'curated by' is a somewhat 'trendy' way of describing 'selected by'. In addition, a separate 'Curated By' role has been proposed to cover books that are catalogues for exhibitions (where a genuine Curator of the exhibition is often named).

There were no other issues raised, and the proposals were ratified.

5. Prioritisation of work on ONIX 3.0.3

LD suggested work by looking at the simpler elements first:

1. renaming of the rarely used <Conference> composite to make it appropriate for other types of event. There were no concerns expressed.

2. addition of <Territory> to composites within <CollateralDetail> to allow market-specific covers or text descriptions (where the product itself is identical, but the marketing of the product is territory-specific). LD related a use case for this in France and French Canada, where one book with one ISBN has two covers or two descriptions in the same language. EB noted that this is seen too in the Netherlands. [Note that a change of cover does not necessarily imply a change of ISBN, though any substantive change such as title, publisher, physical or digital format, or content would). As an aside, AF suggested that in some cases, the book would have the same ISBN in both markets, but use different GTINs. GB asked why this would be necessary, and NG suggested it may possibly be for sales tracking.

JP was concerned for a recommendation by EDItEUR on how to handle multi-territory data where a recipient had no clear composite to choose because none of the composits specified their country – there should be guidance about how to handle a fall-back where the choice of composite is unclear. GB suggested that could be about selecting the first valid composite, or of thinking about language in order of preference. GB agreed such a mechanism should be outlined in the standard (or at least in the best practices). Overall, this proposal was viewed as relatively straightforward.

3. addition of <EpubTechnicalProtection> and <Measure> within <ProductPart>, for more detailed description of multi-component products which are of both print and digital formats. BFS questioned whether there was a real need for this, but AF noted this would be useful for a growing number of 'print + e-book' products, sold under a single ISBN.

4. addition of <EpubLicenceType> into <ContentItem> to deal with hybrid open access. GB said that like item 3, this does happen but not very often. Perhaps some chapters of a monograph are under a CC license and other chapters under a different licence. It is currently very rare – but may become more important in future. BFS said she could see it with custom publishing, but questioned whether it needed an ONIX description? GB replied that custom publishing [eg coursepacks assembled to support a particular college course] would not necessarily need either an ISBN or an ONIX description, but it does sometimes occur. The general view was that the need was too rare at present, so the proposal should be put off until 3.0.4.

At this point EB noted the meeting was running behind, and asked whether we should we should pick particular elements to focus on? GB noted the chair's suggestion that we focus on the simplest in this meeting. However, BFS and LS expressed their need to discuss proposal 5 (one of the more complex proposals).

5. the addition of <EpubTechnicalProtection> and <UsageConstraint> composites into <Price>, to describe a single product that has two different prices, depending on the DRM setup (this would include the use of <UsageConstraint> to describe rentals).

LA explained that in digital, we see that one 'product' can be sold via various business models. Same product, same ISBN, and yet different distribution with separate price or prices for libraries. One price for a small library to lend to 5 patrons, or a different price for a big library to lend to 25. Still the same 'product' and the reading experience is the same. LA noted that according to the existing structure of ONIX, each of these library offers should have a distinct product record and a distinct ISBN. The tags as being used to describe constraints (eg) on lending only exist at the product level and not the price level. His requirement was, within a single ISBN and a single product record, to express a range of different 'commercial offers', each with a distinct price and a distinct range of constraints. Now, ONIX already includes various price conditions and different price type codes, but does not have enough detail to express the full details of a price offer. Proposal 5 is to use the <UsageConstraint> tagging used elsewhere in ONIX at the price level. Currently DRM is not varying alongside this so <EpubTechnical-Protection> is not so important (but may be important in future).

GB noted that the problem revolves around what we mean when we say "the same product". Is the product defined by just the ISBN or by the ISBN and the business model? The ISBN standard currently suggests that significant differences in usage constraints should imply a different product (and a different ISBN). If the product is sold under two different licenses, then it should be considered as two different products – the licence, in many ways, is the product that is sold.

The committee expressed significant dissent. BFS noted that the theory is fine, but the market wants a particular thing here. This is about the complexity of what internal systems can support. This is becoming critical – very critical in Norway. If we don't solve this, then the information stops being in ONIX. BFS has sent the characteristics Bokbasen wants to describe to GB.

EE felt this needed more discussion, to help make the exact requirements clear.

LA commented that he felt the number of business models in digital was expanding significantly. LA further noted that if the price structure included various usage constraints, the next question is what kind of identifier should be used to disambiguate the different commercial offers, eg for ordering, for sales reporting *etc*?

GB noted that <PriceIdentifier> was introduced into ONIX 3.0.2 with exactly this sort of issue in mind. He also explained that in another context – in recorded music – a somewhat similar standard called DDEX treated commercial offers in a way similar to that suggested in proposal 5.

But at the same time, the committee should be wary of creating something that is extremely complex for data recipients (MUCH more complex than to send), and unlikely to be supportable by many current supply chain systems. JP noted a concern for recipients about ordering where placing an order for an item becomes more complicated than merely quoting an ISBN, and this adds a complexity. LA agreed that the proposal is complex, but it will solve a real market need and follow existing ONIX structures in many ways. Perhaps then it prepares ONIX for some evolution.

11. a gender flag in <Contributor>. GB explained that this is a suggested data element for ISNI registrations, and would make registrations based on a data in an ONIX product record more likely. He added that while this proposal had caused a certain amount of consternation at the UK and US national group meetings at which it was raised, the level of concern was greatly reduced

after a little explanation. The flag would specify the gender (not the much more complex sexuality or sexual identity) of a persona (the outward public face of a contributor, not of an actual person). As example, the gender of Robert Galbraith and of George Eliot would be male – even though the real people behind those names are both female. It would not be mandatory in any way.

JP noted that this information was already present in many library authority files, and in ISNI metadata. FF expressed support for the proposal, AF disputed the need.

Given the time constraints of the meeting, the committee agreed that further discussion – of proposal 5 and the other proposals – should be devolved to a small technical subgroup, which should develop a recommendation for ratification by the committee. GB agreed, and FF, LA, EB, KL, MR volunteered. LA suggested the subgroup should aim for regular weekly 1hr calls.

ACTION: GB to convene the subgroup and arrange conference calls.

UPDATE: GB has also asked Tom Richardson of the Canadian national group to join the subgroup

GB asked the committee whether the subgroup should aim for publication of 3.0.3 in January as intended. FC suggested the achievability of the target itself should be the first item on the agenda for the subgroup.

GB reminded the committee that these proposals were not suitable for discussion on the ONIX_implement mailing list, and that a closed group (similar to the *Thema* closed group) would be set up for the subgroup and committee discussions on this.

6. Any Other Business

There was no other business.

7. **Next Meeting**

LD noted that we had run out of time, thanked participants, and advised the next meeting would be at the London Book Fair, on the afternoon of Wednesday 13th April 2016 at a location to be confirmed.

Alex Ingram / Graham Bell
EDItEUR
15th October 2015