Executive Director's Report to Members: October 2012

1   Overview

The 6 months since the last Open Meeting seem to have passed in a flash; indeed the same is true of the 12 months since the last Frankfurt Book Fair. As I write, the EDItEUR office is in the final throes of preparing for the string of EDItEUR meetings in Frankfurt – ICEDIS; the Frankfurt Tools of Change *Supply Chain Track*; our accessibility seminar: *A New Market: accessible ebooks in mainstream channels*, the ONIX for Books International Steering Committee; our Open Members Meeting; and our Board meeting on October 11th.

In terms of EDItEUR's own standards, the last 6 months have been ones characterised by consolidation and the continuing encouragement of implementation. There are fewer obvious milestones to note in terms of new specifications published; but this is not an indication that we have been sitting back. The office remains as busy as ever.

2   Membership

We have been delighted to welcome the following new members during 2012:

- Batch
- Biblioteka Narodowa (the Polish National Library) (Associate Member)
- Capita
- Frankfurt Book Fair
- GiantChair
- Ithaka/Portico
- Kogan Page
- SAGE
- Türkiye Yayincilar Birligi (the Turkish Publishers' Association)
- Waterstone's

We are planning another membership campaign for 2013.

3   EDItEUR Board news

Paola Mazzucchi (AIE) takes the chair for the first time at our Board meeting in Frankfurt.

We are grateful to her and to the other members of the Executive Committee – Len Vlahos (BISG), Vice Chair and Neil Wilson (British Library), Treasurer – for their support.

4 EDItEUR Standards development

4.1 ONIX for Books

4.1.1 Codelists

Issues 17 and 18 of the ONIX for Books codelists have been released since the last Open Meeting. Highlights of these issues are:

- various new codes for identifiers and classifications used in the Chinese and Japanese book trades – for example the Chinese OLCC number as a product identifier in List 5 and a Japanese children's audience code in List 29 and codes for common Chinese ebook formats;
- a new code value for the German National Library's Gemeinsame Normdatei, a combined authority file and identifier scheme that's replacing the separate Schlagwortnormdatei, Gemeinsame Körperschaftsdatei and the Personennamendatei authority files for subject headings, personal and corporate names (which are deprecated);
- new date and time formats to support timed releases of books and e-books (eg embargoed until 11:15am on a particular morning)
- an entirely new List 203. The intention of List 203 – which can be used with ONIX 2.1 or 3.0 – is to allow publishers to attach a broad content warning to products that are suitable for an Adult audience, but that might, in other sectors, attract 18+, X or similar restrictions, audience ratings or warnings. It can indicate, for example, that a romance novel contains passages of an explicit nature, or that a thriller is particularly violent. Over and above ordinary subject classification, this information can be extremely useful to retail merchandisers, and to librarians offering automated selections to their borrowers.

Issue 19 is in preparation, and the very short list of suggested new codes has been circulated to ONIX National Groups for comments. It is expected that issue 19 will be approved at the Frankfurt Book Fair ISC meeting and will be released at the end of October.

A new avenue we will explore over the next few months is the publication of the ONIX codelists as SKOS, so they are suitable for use with Linked Data (semantic web) initiatives in the library sector.

4.1.2 Enhanced schema for ONIX

ONIX implementers benefit from a range of tools to check the correctness of ONIX XML messages. The ONIX DTD, or the XSD and RNG schemas can be used to 'validate' an ONIX message. The DTD checks only the syntactic structure of the XML – are the right tags used, in the right order, and are they nested properly? The XSD and RNG schema files validate the structure, but also check some of the data content – is this code taken from the right codelist, is that number an integer, is that text too long? But the checks an XSD or RNG schema can implement are limited, and the ONIX schemas deliberately don't even do all they could – they don't enforce maximum lengths on text data, for example, since the standard merely recommends maximum lengths.

Codelist releases 17–19 have also been accompanied by slightly updated schema files (XSD and RNG), which do not change the standard but do check conformance a little more rigorously. Continued small additions and clarifications have also been made to the ONIX 3.0 Implementation and Best Practice documentation as issues have been raised (eg on the ONIX_implement mailing list).

Going beyond XSD validation, Schematron is a type of rule-based enhanced schema that can check the deeper structure of an ONIX file. It can validate the check digit of an ISBN algorithmically, or check co-occurrence constraints. There are many of these in ONIX – where data in one element affects the validity of another, for example, an e-book cannot have physical measurements in <Measure>, or if an identifier for a contributor is proprietary (eg a publisher's own internal author number), it must be accompanied by an <IDTypeName> element.

And as well as flagging such errors, it can produce warnings – for example if a text element like a short description is too long, or if the ONIX does not follow some particular best practice.

A full Schematron validation tool for ONIX 3.0 is in preparation: the first (incomplete) version will be released on the EDItEUR website before the Frankfurt Book Fair. It's not a tool for everyone, but for developers it will provide a way of evaluating ONIX messages in an automated way.

A small number of organizations which receive large volumes of ONIX data have expressed an interest in Schematron as a way of pre-processing and accepting/rejecting messages prior to ingestion into their production systems. But it also helps to further codify the ONIX structure and best practices, and development of the Schematron has already led to a handful of clarifications in the Best Practice documentation and the main schema files.

4.1.3   ONIX for Books training

On behalf of Book Industry Communication (BIC), Graham Bell delivered another successful ONIX for Books training session in London in May. Once again the course was fully subscribed, and attracted both beginners and experienced delegates from all parts of the supply chain. Graham is repeating the course for BIC in November.

During a busy late June week in Italy, Graham delivered an in-depth training session in Milan for staff from AIE and Informazione Editoriale, and DOI registration agency mEDRA. He also delivered a webcast focusing specifically on differences between ONIX 2.1 and 3.0 that formed part of BISG's Evolving Metadata Practices series. A recording of the BISG webcast can be downloaded from http://bit.ly/LHwlI6.

4.2   ONIX-PL

Issue 4 of the ONIX-PL Dictionary was published in June on the EDItEUR website at http://bit.ly/NYPrHi. Some thirty-nine new terms have been added, almost all in response to specific user requests. The resources now available include updated versions of the ONIX-PL Schema files, a full Dictionary spreadsheet listing all terms and their definitions, and the Dictionary file as used in the OPLE editing software (with new instructions for installing the OPLE version). As usual, we have also provided a separate summary list of the newly-added terms.

4.3   ICEDIS

During the last 6 months, we have continued to promote and advocate the implementation of two ONIX formats recently introduced. ONIX-PC (the ONIX Price Catalog message for subscriptions products) was again generated by a number of leading publishers over Summer 2012 to convey details of 2013 subscription pricing. Several initiatives are underway to broaden the rollout of ONIX-PC usage during 2013: these include proposals for a targeted drive within the ICEDIS membership, improved documentation and sample files, and focused training to assist implementers.

Meanwhile, ONIX-PH (ONIX for Preservation Holdings) is under review by agencies contributing to the Keepers Registry, hosted by EDINA, with the aim of eventually endorsing this format as the vehicle of choice for communicating complex holdings statements about preserved e-journals. ONIX-PH is itself a dedicated subset of the earlier ONIX for Serials SOH (Serials Online Holdings) format

4.3.1   ICEDIS at NASIG: headline topics

In early June, we held an ICEDIS Committee Meeting in Nashville, TN, USA, during the NASIG conference. Our NASIG friends made us most welcome and we were able to brief North American participants about recent EDItEUR/ICEDIS achievements and project work streams that ICEDIS has underway. The meeting was attended by a number of long-time EDItEUR members, together with representatives from several organizations that wanted to know more and are considering membership.

4.3.2   A new Co-Chair for ICEDIS

One important piece of business at the ICEDIS Meeting in Frankfurt will be the election of a new Co-Chair of the Committee, to replace Ramon Schrama of Swets who will be standing down after seven years in the post. We will be thanking Ramon for his excellent contributions and dedicated guidance of ICEDIS over such a long period, as well as welcoming a newcomer to the "hot seat"! The other Co-Chair of ICEDIS – Keith Whiter of Elsevier – will continue in office.

4.3.3   Structured Name & Address Model

A very active ICEDIS Working Group has made significant progress over the past year in developing a standardized model to communicate structured names, addresses and party identifiers. The model is expressed in XML and draws on many concepts and data elements familiar from the ONIX range of standards. We have now reached a point where the draft model seems sufficiently robust that we can publish an overview description on the EDItEUR website – both to communicate the work to date and to invite wider suggestions and critique. We are delighted to report that this model is now available and can be found on the EDItEUR website at http://bit.ly/PzlJsT.

4.3.4   Legacy standards and changing business needs

Recently there have been several suggestions to look again at some of the legacy standards that EDItEUR maintains, notably the ICEDIS fixed-format standards and some EDIFACT standards used for serials. Quite naturally, some of these formats have been overtaken by the emergence of business models that didn't exist when the formats were designed. But there are clearly trade-offs to be struck when considering whether or not to amend established standards that have been so widely implemented. We will debate the pros and cons of these suggestions during the Frankfurt ICEDIS Meeting and try to decide upon the most effective way forward.

4.3.5   Working Group updates

In addition to the Structured Name & Address model (see above), other Groups have been collaborating to develop a standard for New Subscription Orders and to devise best-practice recommendations for minimizing the pain associated with Price/Payment Reconciliation. Representatives of both of these groups will be reporting back on progress and asking for guidance from the ICEDIS Committee on how best to bring the projects to completion.

The New Order work has a particular significance, since it both speaks to a clearly defined business need and represents essentially a prototype for future message development in the subscription transaction area. Working drafts of both the New Order message and the Price/Payment Reconciliation best practice recommendations will be published during the next month.

4.3.6   Publication of ICEDIS specifications

We have been rationalising the presentation of our serials/subscriptions standards on the EDItEUR website. A new category has been introduced in the Standards section, entitled ONIX for Subscription Products. This has allowed us to group together the developing range of standards relevant to the subscription supply chain, including both newer entrants like ONIX-PC, ONIX-PH and the Structured Name & Address model as well as the longer-established ONIX for Serials formats. In a similar way, another new category, "ICEDIS Fixed-length Formats" now signposts the legacy standards for serials subscriptions and renewals that are still in very widespread use across the industry.

4.4   EDItX

With revenue from ebooks now increasingly important to publishers and their supply chain partners, there have been moves in several countries to standardise the way that ebook sales are reported through the supply chain, and the EDItX Sales Report Version 1.1 standard is proving to be what is needed. In Germany publishers, intermediaries and retailers have recently agreed a sales reporting specification derived from (that is, a subset of) the EDItX standard. The specification is available (in German only) from http://bit.ly/PSq3p0.

Since its publication just over a year ago, in parallel with the EDItX standard of which it is a subset, the Book Industry Study Group (BISG) sales reporting standard has been implemented by a number of major publishers and other key players in North America. BISG has just announced plans to review the standard and make revisions to reflect how reporting requirements have evolved over the past year.

In North America and in the UK there are signs of interest in extending the usefulness of the sales report format to enable its use for reporting revenue streams from sources other than conventional ebook sales. Both Book Industry Communication (BIC) in the UK and BISG are exploring with their members whether the existing format should be extended to enable its use in reporting revenue streams such as subscriptions and rentals.

5   Identifier standards for which EDItEUR provides management services

5.1   International ISBN Agency

EDItEUR has recently renewed its management contract with the International ISBN Agency for another three years. The agency hosted its 40th Annual General Meeting in September; this year's meeting took place in Chisinau, Moldova and was well-attended, with delegates from around 40 national ISBN agencies across the world participating. It was an extremely well-organised event, thanks to the considerable efforts of the National Book Chamber of Moldova (which among its many responsibilities manages the Moldovan ISBN Agency). Among the topics discussed during the AGM,

there was a discussion about the ISO processes for revisions of the standard (the ISBN comes up for "systematic review" in 2013).

5.2   International ISTC Agency

We have recently renewed our contract for providing administration services to the International ISTC Agency for a second year. This is a low level administrative role in comparison with our major commitment to ISBN.

5.3   International ISNI Agency

We have also signed a contract to provide the same administration services for ISNI as we provide for ISTC.

6   Major projects

6.1   ARROW PLUS

The ARROW PLUS project is an EC-supported "Best Practice Network", which builds on and further implements the ARROW system. EDItEUR's specific role in the project – led by Brian Green – has been to assist in the development of Books in Print services (BiPs) in a number of countries without them. Our first task was to build a common set of Business Requirements. Brian is now assisting the countries which are moving ahead with creating BiPs to develop business plans with a view to making them sustainable once project funding ceases.

6.2   Linked Heritage

Linked Heritage (www.linkedheritage.eu) is a 30-month, European Commission funded project aiming to contribute metadata for objects of cultural heritage interest for discovery in web portals. It centres on the EC's flagship portal, Europeana (www.europeana.eu). EDItEUR's contribution is to identify the commercial sector's metadata standards and best practice, the technical means of transforming and integrating commercial data within existing heritage portals, and finally the legal and licensing agreements that might encourage commercial companies to share their data in return for links to buy the described items.

As the most recent part of its contribution to the project, EDItEUR and its project partners have produced a comprehensive mapping of ONIX for Books 3.0.1 to the LIDO cultural heritage metadata schema (http://bit.ly/L8A0sg), implemented in XSLT. This should allow publishers to contribute book data to Europeana where, alongside relevant museum, library and art gallery objects, the books will appear in search results, with a Web link to allow browsing to the publisher's site for purchase.

The ONIX for Books mapping is the central result in a report on the interoperability of commercial product data with the heritage sector's standard data models, also taking in film and TV data (EIDR and ISAN standards), photos (IPTC and XMP) and recorded music (DDEX). The report builds on EDItEUR's first deliverable for Linked Heritage, a best-practice report surveying the commercial standards (available at http://bit.ly/S5Ec3P), and contains detailed recommendations for pushing forward semantic interoperability across the media sectors and between commercial creative industries and cultural memory institutions. MVB has worked on a similar ONIX 2.1 to LIDO mapping.

Our next (and final) tasks in this project concern the legal and licensing framework within which commercial organizations might contribute their metadata to Europeana.

6.3   DCH-RP

We have recently signed a contract for our participation in *Digital Cultural Heritage – Roadmap for Preservation*. Our work on the project will be led by Tim Devenport. We will have more to report in 2013.

6.4   Enabling Technologies Framework (accessibility)

We have recently renewed our contract with WIPO for the third and final year of this project. Our major productive task this year is to continue to build on the online training course, the first module of which is already launched (http://bit.ly/MxTDdg). Members who do not know much about accessibility issues might find this module (which is expected to take about 30 minutes to complete) useful. We expect there to be 6 modules by the middle of next year when the project finishes.

Sarah and I are also pursuing a particular line of work as our contribution to the UK's Accessibility Action Group (chaired by the PA). This is to explore the accessibility of library ebook platforms (particularly those in use in academic libraries). We are working with JISC TechDis on this project.

To contextualise this activity, in April, just before the London Book Fair, the UK Publishers Association published a Joint Statement on Accessibility and eBooks on its website (http://bit.ly/JlaZIT). This statement (which was initially drafted by Richard Orme of the RNIB and myself) was created jointly by the Publishers Association, three of the UK's advocacy organisations working on behalf of people with reading impairment – the RNIB, Dyslexia Action, and JISC TechDis – and EDItEUR. The statement commits its signatories to work to make all e-books as accessible as possible at no additional cost to their users. By following EDItEUR's guidelines , publishers can work towards ensuring that all their e-books have easily-added features that assist accessibility (like enhanced navigation); and equally that their e-books do not have features that aid accessibility disabled unnecessarily. However, publishers can only get so far in improving the accessibility of e-books – the creation of accessible digital files is only the first stage of the answer. Each part of the supply chain – particularly the developers of e-book platforms and devices, but also distributors, retailers and libraries – has its own part to play. Print-impaired readers themselves also have a contribution to make through gaining an understanding of which accessibility features and assistive technologies are best suited to their individual needs.

EDItEUR has also undertaken some limited work for the TIGAR project (Trusted Intermediary Global Access Repository), a separate WIPO project that aims to ease the exchange of accessible electronic versions of books between libraries for print-impaired readers. This work aided development of the TIGAR metadata model, central repository and its user interface. It also ensured that although the metadata is MARC-based, it can be easily modified to accept ONIX when TIGAR moves into its next phase and begins to involve commercial publishers.

Graham Bell attended part of the annual IFLA World Library and Information Congress (http://bit.ly/OErhFF) in Helsinki in August, and presented to IFLA's *Libraries Serving Persons with Print Disabilities* section on how new e-book formats like EPUB 3 and improved metadata coverage of accessibility options can help make ordinary, commercial e-books suitable for blind, partially-sighted or otherwise print-impaired readers.

7   Beyond EDItEUR

7.1   EPUB 3

EPUB 3 is unsurprisingly gaining considerable traction in becoming the standard for ebook content delivery. We hear that major publishers are now testing EPUB 3 files in the supply chain. There is particular enthusiasm for its capability in managing much more complex layout requirements than are possible in EPUB 2 (or indeed any other reflowable format); and also for its in-built capability in managing interactivity. The integration of the DAISY delivery format into EPUB 3 is also of course very significant in the context of our "mainstreaming accessibility" agenda.

For members of EDItEUR who may not have seen it, BISG has created an extremely valuable tool comparing the EPUB 3 capabilities of different ebook platforms. The BISG EPUB Grid provides detailed information on what enhanced functionality available in the IDPF's recently released EPUB 3.0 standard is supported by which devices, applications, and reading systems. The Grid is freely available from http://bit.ly/M7pTJN . Recognising the rapid changes in digital book publishing and retailing, BISG is reviewing and updating the Grid continuously. Industry feedback plays a key role in the update cycle; feedback is invited to info@bisg.org.

7.2   Linked Content Coalition

The LCC project continues, although it is staying rather "under the radar" at the moment. My limited *pro bono* role continues to be to lead the advisory group on Governance. Work is continuing on developing what is likely to be LCC's major deliverable, the Rights Reference Model (RRM) and its XML representation, the Common Rights Format.

7.3   The UK "Copyright Hub"

This is the potential realisation of the so-called *Digital Copyright Exchange* proposed by the Hargreaves report. The proposal appears to be getting fairly wide-spread cross-media support. An encouraging aspect of Richard Hooper's feasibility study into the DCE – which has led to the development of the Copyright Hub – was its close focus on the critical importance of identifiers and metadata.

7.4   BISG metadata report

BISG has published the results of its study commissioned from Magellan Media Consultants entitled The Development, Use and Modification of Book Product Metadata. The study mapped the flow of product metadata from publishers to eventual users of the data, and the issues that arise as it passes through the metadata supply chain, based on a survey of around 125 key players in the US book trade. It endorses the hands-off exchange of data characterized by use of ONIX, encourages update to ONIX 3.0, and above all emphasizes the close relationship between business process improvements and increased accuracy of metadata fed into the supply chain. Further details: http://bit.ly/NlvWYS.

MB – 3 October 2012