An Introduction to ONIX for Books

What is ONIX?

ONIX – or more precisely ONIX for Books – is a framework encompassing standardised terminology, operating methods and a data file format that's used to communicate information about books, audiobooks, e-books and other related products. The data format is widely used across the global book publishing industry to pass data between publishers, data aggregators, wholesalers and distributors, and High Street and web-based retailers.

The most visible part of this framework, the ONIX file format, is a standardised data file that describes a publisher's products – the bibliographic details including critical information such as ISBN, title, contributors, subject, publication date and so on. But it also encompasses marketing collateral for the products, including descriptive text, reviews, the cover and other promotional images. Finally, ONIX includes commercial information such as suppliers, prices and sales rights. All this information – 'metadata' – is arranged in a carefully structured XML-based message format that is sent from a 'data provider' (typically a publisher or data aggregator) to a 'data recipient (typically an aggregator which then re-sends it elsewhere, or a distributor, wholesaler or retailer).

Allied to this, ONIX provides a framework of terminology, data structures and conventions that are widely used within publishers' IT systems and those of their service providers and retailer customers.

The benefits of ONIX

Comprehensive, accurate and timely product information is critical to 'discoverability' by end-customers online, and drives effective merchandising in traditional High Street stores.

Retailers depend on the accuracy and timeliness of product information available from publishers. They demand rich, reliable and up-to-date metadata on which to base their back office, store systems or online catalogues.

The ONIX standard was created to meet the need within the book supply chain for communication of this rich and comprehensive yet unambiguous and machine-readable product information. It allows publishers and supply chain partners to automate the management, distribution, updating, processing and display of a substantial volume of metadata about hundreds

of thousands of products, thereby removing cost from the supply chain. It also improves the availability, timeliness and quality of the metadata that the supply chain is dependent upon, in turn aiming to ensure that sales are not lost through missing or inaccurate information.

And ONIX enables publishers to collate this data into a single definitive product catalogue which can be distributed in the same standard data format to all retailers, data aggregators, affiliate companies and distribution organisations. Since ONIX is a global standard, the scope of this is international: the same ONIX data sent to a UK data aggregator can in principle also be sent to a US-based retailer or a Chinese e-book vendor.

Many publishers have also extended the potential of ONIX by using their records to create traditional advance information sheets and catalogues based on the same core data, and by using ONIX data as the basis of their own websites. The expectations that ONIX carries in terms of the structure and content of a 'product record', and in terms of operating practice, have also been of great benefit, ensuring consistency and data discipline across publishers of different types. This has spurred a competitive market for IT vendors delivering off-the-shelf software solutions, and enabled data aggregators to provide a more comprehensive and more valuable service to the sector as a whole.

ONIX is an open standard, developed and maintained by EDItEUR, the international book industry standards organisation, along with an international steering committee representing the supply chains in countries where ONIX has been widely adopted. The committee includes delegates from countries spanning North and South America, Europe and the Asia-Pacific region. Book Industry Communication (BIC) provides representation from the UK, although EDItEUR itself is also based in London. This governance structure ensures that ONIX continues to develop to meet new commercial challenges in the market, while at the same time ensuring it benefits the supply chain as a whole and not just one narrow commercial interest.

A little history

ONIX was first outlined in January 2000, as guidelines arising from a working group convened by the American Association of Publishers. It became a practical standard a few months later, as EDItEUR released ONIX version 1.0. Several subsequent minor revisions added new features, yet until the release of version 2.0 in mid-2001 there were only a handful of early adopters in the UK and USA.

ONIX is built using XML, a simple and generalised mark-up syntax developed by the World Wide Web Consortium, using 'tags' enclosed in < and > characters to define data elements. XML is widely supported by various software tools, and is used in many publishing applications other than ONIX.

With the advent of ONIX version 2.0, the number of data providers and recipients grew quickly, spreading to Germany, Spain and beyond. Version 2.1 – dating from 2003 – accelerated that trend. There are now thousands of data providers and hundreds of recipients.

The explosion of the market for e-books that began around 2008, and the increasingly international nature of the book business, prompted development of ONIX 3.0, which was released in late 2009. This adds new functionality and refinement, but at the same time has a simpler, more logical yet more flexible structure. It is not entirely backwards compatible with 2.1.

Early versions of ONIX have been obsolete for many years. Version 2.1 has been frozen since the development of 3.0 began – the last update for use in the UK was in 2006 – and most EDItEUR support for 2.1 ended in 2014. However, version 2.1 is still in use by a significant proportion of data senders and recipients in the UK and some other countries, and archived documentation is still available.

EDItEUR's policy is to review the business requirements for ONIX regularly, and small evolutions of the standard are published every two years or so. Hence version 3.0 has become version 3.0.5, adding new functionality while retaining full backwards compatibility with the earlier releases of 3.0. Codelists – an important part of the framework – are also regularly revised, with Issue 44 released in January 2019.

New adopters of ONIX are expected to implement ONIX 3.0 from the outset, and key UK data senders and recipients are able to make use of 3.0. The older version 2.1 is likely to be supported by most data senders and recipients in the UK supply chain for a few more years, but the business risk of relying solely on version 2.1 is gradually increasing. All existing ONIX users are recommended to adopt 3.0 quickly.

Who uses ONIX?

Almost all the major book publishers in the UK use ONIX to transmit their product information to metadata aggregators including Nielsen, Bowker and BDS. They in turn re-distribute it in aggregated form, as ONIX or another format, to many more 'downstream' organisations (distribution and wholesale companies, retailers, libraries etc). Most publishers also use ONIX to communicate directly with other data partners – service providers, distributors, wholesalers, e-book vendors and even some individual High Street and online retailers. It can also form a part of the legal deposit process and of publishers' communications with PLS.

In total, data aggregators in the UK receive ONIX from around 500 data providers, from the largest publishing groups to small independents, and this number continues to grow. Together, these 'metadata feeds' represent an overwhelming proportion of all the books sold in the UK.

ONIX is equally widely used in many other European counties and in North America, with a growing number of implementations in Latin America and the Asia-Pacific region (including China and Japan).

How to start with ONIX

ONIX as a standard data format should be viewed first as a 'side effect' of good data management practice. For a publisher or retailer, committing to ONIX is really a commitment to the improvement of metadata quality and management processes across the business. Efficient exchange of this data with external partners in the ONIX data format is only a part of this.

However, adoption of the standard is clearly a challenge. ONIX is by its nature a complex data standard and it might initially appear beyond the capability of small and mid-size publishers and retailers. Yet it is certainly possible for even the smallest publisher to start from scratch and build a bespoke system using only the Specification document and Implementation and Best Practice Guidelines which are published on the EDItEUR website – and several have done so successfully. However, development of this sort requires internal technical expertise that only a few small publishers possess. Most small, mid-size and even many large publishers choose to invest in an off-the-shelf IT solution that provides comprehensive product data management capabilities. Almost all such systems on the market are capable of producing ONIX files in line with the standard. The third approach is to take an existing in-house system and add ONIX communication through custom development work. Whichever approach is taken – an entirely in-house development, an off-the-shelf solution or custom additions to an existing system – cultural changes within the organisation will likely also be required, to ensure clear ownership of the metadata management process, adequate resourcing and high-level management support. High-quality metadata is valuable. Too often, metadata is seen as 'something for the intern', and neither a good IT solution nor ONIX itself can mend inadequate or conflicting data.

Off-the-shelf product management systems suitable for publishers from the largest to the very small are available from BooksoniX, Consonance, Focus, GiantChair, Ingenta, Klopotek, KNK, Stison, Trilogy, Virtusales and many others. Almost all of these are ONIX 3.0 capable, and there is a mix of on-premises and hosted solutions.

For retailers, there's no ready-made choice of off-the-shelf applications, though there are some solutions like Supafolio (a Wordpress plug-in), Shopify plug-ins and Ingram's aerio that are worth exploring for small-scale online retailing. In the UK, wholesalers provide online ordering systems to retailers which make use of ONIX data in the background, and Gardners' Hive can help integrate online and High Street retailing. And the largest retailers maintain their own in-house product catalogues, either through direct receipt of ONIX files or through non-ONIX data feeds from aggregators such as Nielsen BookData or BDS. The fact that ONIX is based on XML

means there are tools available for parsing the data and inserting it into most database management systems.

Finding out about ONIX

The definitive source of information about ONIX is EDItEUR. Its website (https://www.editeur.org) gives access to four key groups of resources:

- The latest release of the *Specification* for the data format (this is currently version 3.0.5, though its generally just known as ONIX 3.0). There is specification documentation and XML tools (primarily XSD schemas and DTDs) for developers that together define the data format and ensure that it is correctly constructed
- The latest *ONIX 3.0 Implementation and Best Practice Guide*, which provides practical interpretation of the *Specification* and many real-world examples of how the ONIX data format is used. This forms the global basis of 'good ONIX practice', though individual countries might also make a few additional recommendations. EDItEUR also posts detailed 'application notes' and other guidance documents
- The latest release of the *ONIX codelists*. These are codes used in conjunction with the *Specification* – in some ways they are the nouns in the ONIX 'language'. There's also an interactive online browser for the codes at https://ns.editeur.org/onix
- Documentation and XML tools for older versions of ONIX, including ONIX 2.1.

The EDItEUR website is also the source of information on *Thema*, the subject classification scheme that can be used with ONIX (or independently), and EDItX, a range of inventory and sales report formats and XML-based transactional messages. You will also find links to a number of other ONIX-branded messages which have been created to provide different functionality, for example for journal publishers. These other ONIX messages share much of the character of ONIX for Books, but are intended for use in other parts of the wider publishing industry.

Community-based advice and support is also available through a mailing list moderated by EDItEUR https://groups.yahoo.com/neo/groups/onix_implement/info

All of the above resources are freely available, and ONIX itself can be implemented free of charge under a permissive licence from EDItEUR (https://doi.org/10.4400/nwgj). Both BIC and EDItEUR are membership organisations, and depend on member revenue to make this possible.

Establishing an ONIX feed

As a data provider, once you have ONIX capability in your chosen product management system, there are several steps you need to take before you can start transmitting product information.

The first is to make sure the data you hold internally is clean and accurate – this task should not be underestimated. You'll have to establish who in your organisation is responsible for creating and maintaining the data, and who will be allowed editing rights. Establishing clear processes, roles and responsibilities for data management is vital if your ONIX capability is to deliver any business benefit. It is a mistake to view adoption of ONIX as an IT project.

The second is to decide which elements of your system – and of ONIX – you are going to use, both internally and externally. ONIX is comprehensive, much more comprehensive than the requirements of any single publisher or retailer, and many parts of ONIX are optional. Not all of it will be appropriate to your business. And IT vendors build systems that manage data – for example production scheduling or digital asset management – that has no place in ONIX.

For the UK book trade, Nielsen BookData has published guidelines for a practical minimum set of elements which are expected in any ONIX feed, and which can then be communicated in outbound feeds to booksellers and libraries (see http://www.nielsenbookdata.co.uk/uploads/BookData ONIX Guidelines V3_2.pdf). However, this document pre-dates ONIX 3.0 and while many principles apply equally to 2.1 and 3.0, the details are not all up to date. EDItEUR's ONIX 3.0 Specification contains an annotated 'example product record' that can be used as a guideline. However, the 'minimum set of metadata fields' for a publisher or retailer is dependent on the nature of their business, and requires thorough analysis.

The third step is to contact data partners and aggregators and tell them that you will have ONIX files available. Your files will most likely require testing and validation by potential recipients, before they accept that your feeds can be ingested automatically into their systems.

Equally, for data recipients, establishing a feed with a new data supplier requires an analysis of your particular business and data requirements, then based on this, agreeing a minimum set of ONIX data elements and expected levels of service with data providers. Testing the ingestion of data in an initial trial data feed is essential, as future data provision won't be 'curated' – updates will be largely automatic, and that updated data could be presented to business customers or consumers without any manual intervention. It is important that recipients are able to accept data feeds that contain more data than the minimum, even if some is then largely ignored.

Publisher's products have a long lifecycle, beginning months or even years before publication and extending to or beyond the point at which the product is declared out of print. During this time, details of the product change – most obviously, the price and availability are dynamic, but in the months leading up to publication, almost any data element is subject to change.

So it's important that once a data feed is established, data providers and recipients update their metadata records whenever changes are made to the product information by the publisher. This means establishing a routine to send or receive updated ONIX records alongside brand new records (in fact, in most data feeds, replacement records with updated information predominate.)

Post-publication, the set of data fields subject change is limited. In part because of this, ONIX 3.0 updates can be 'partial updates' – that is, an update may contain only a partial product record. Most typically, marketing collateral (Block 2) and price and availability (Block 6) might be updated without including any other data in the record. This partial or 'Block update' mechanism is more efficient than sending whole records when only a small part of that record has changed. Alternatively, some data providers and recipients take the opportunity to use another data format for post-publication updates (for example, if price and availability are the only post-publication updates, then an EDI format such as Tradacoms will suffice). However, this information is critical to customers' ability to buy the books and, whatever process you adopt, close attention should still be paid to its accuracy – and particularly to the need to keep both types of data feed 'in sync'.

Accreditation

BIC manages an accreditation scheme which allows publishers and other data providers to demonstrate that they supply comprehensive, accurate and timely metadata to the supply chain. The BIC Product Data Excellence Award rewards data providers delivering at least a defined minimum set of elements, within strict timing guidelines. At present, the Gold and Silver level awards are only available to ONIX users, and in the near future, a revision of the scheme will require ONIX 3.0 to gain Gold level accreditation (and adoption of many ONIX codes and field definitions even for the most basic Bronze level). Full details of the current scheme and those companies which hold the awards can be found at http://www.bic.org.uk/90/Product-Data-Excellence-Awards/

ONIX in the UK

ONIX for Books is supported in the UK by BIC (Book Industry Communication) and its UK ONIX National Group, which provides input to the international development of the standard and promotes use of ONIX in the UK market. BIC also has direct access to the EDItEUR staff and consultants who have been the architects of the standard since its inception, and who can provide authoritative answers and advice.

EDItEUR and BIC also provide face-to-face training on ONIX on a commercial basis, with regular ONIX: essentials and ONIX: advanced topics courses held either in central London or in-house at EDItEUR or BIC members' premises. There are also occasional workshops on specific aspects of ONIX practice.

Contacts and resources

Book Industry Communication
7 Ridgmount Street
London WC1E 7AE
T: 020 7255 0516/3
E: info@bic.org.uk
www.bic.org.uk

Nielsen BookData
Midas House
62 Goldsworth Road
Woking
Surrey GU21 6LQ
T: 01483 712200
E: help.book@nielsen.com
www.nielsenbookdata.co.uk

Bibliographic Data Services (BDS)
Annandale House
The Crichton
Bankend Road
Dumfries DG1 4TA
T: 01387 702251
E: info@bdslive.com
www.bdslive.com

Bowker (UK)
3 Dorset Rise
London EC4Y 8EN
T: 020 7832 1700
E: datasubmission@bowker.co.uk
www.bowker.co.uk

EDItEUR
United House
North Road
London N7 9DP
T: 020 7503 6418
E: info@editeur.org

———————

BooksoniX
The Milking Parlour
Hurst Farm, Dairy Lane
Crockham Hill
Kent TN8 6RA
T: 01883 343000
E: contact@booksonix.com
www.booksonix.info

Consonance
112 High Street
Thame
Oxfordshire OX9 3DZ
T: N/A
E: info@consonance.app
www.consonance.app

Focus IT Services Limited
Suite 108 Regent Business Centre
291 Kirkdale Sydenham
London SE26 4QD
T: 020 8776 5979
E: info@focusservices.co.uk
www.focusservices.co.uk

GiantChair (onixsuite)
1875 Connecticut Avenue NW
Washington DC 20009
USA
T: N/A
E: www.onixsuite.com/contact
www.onixsuite.com

Ingenta (Vista)
8100 Alec Issigonis Way
Oxford, OX4 2HU
T: 01865 397800
E: info@ingenta.com
www.ingenta.com

Klopotek
Luminous House
300 South Row
Milton Keynes MK9 2FR
T: +49 30 884 53 244
E: A.Lindsay@klopotek.co.uk
www.klopotek.co.uk

KNK Publishing Software
71-75 Shelton Street
London WC2H 9JQ
T: 01707 828020
E: info@knk.com
www.knkpublishingsoftware.com

Stison Publishing Solutions
2 Church Row
Wandsworth Plain, Wandsworth
London SW18 1ES
T: 020 7183 1476
E: info@stison.com
www.stison.com

Trilogy Enterprise Services
Aries House, 43 Selkirk Street
Cheltenham
Gloucestershire GL5 2HJ
T: 01242 222132
E: enquiries@trilogygroup.com
www.trilogygroup.com

Virtusales Publishing Solutions
Sheridan House
112-116 Western Road
Hove BN3 1DD
T: 01273 917373
E: info@virtusales.com
www.virtusales.com