



Data Recipient Best Practice

Project Glossary

Prepared by Stephen Long
October 2024 Version 1.0

Copyright © 2024 Book Industry Communication Ltd.



Disclaimer:

Please note: The information provided in this document is intended for guidance purposes only. Those involved in the creation, collection, management or distribution of product metadata are strongly advised to seek guidance on compliance with the business policies of their respective organisations.

GLOSSARY:

With particular thanks to Booksolve, EDItEUR, Rakuten Kobo and Taylor & Francis for their contributions to this glossary.

Key terms that are themselves defined in this glossary are underlined.

Abridged: Content shortened by removal of text and minimal re-writing, occasionally termed condensed (the latter implies a greater degree of re-writing).

Academic and Scholarly: Publishing, library and bookselling sector concerned mostly with the university market, academic research and scholarship.

Accessibility: The practice of making publications usable for as many people as possible.

Adapted: Content that has been modified to create a new work, often to make it suitable for a new medium (i.e., book to film).

AI or Artificial Intelligence: Also known as “machine intelligence”. Broad term for the application of software to analyse data and make decisions based on that data. Aimed at achieving some predefined goal. Encompasses natural language processing, knowledge representation, reasoning, machine learning etc. In publishing and metadata, AI techniques might for example be applied to automated entity extraction (recognising names, places, concepts and so on) from the text of a book to create keyword lists, indexes or links to other books about the same entities, to sales prediction, to subject coding and sentiment analysis, to abstracting and summarisation of the text of a book, or to generating new text or images (from a marketing description or single illustration to the content of a whole new book). The last application – termed generative AI – is controversial due to the use of copyright material as training data.

AS2, AS3: Applicability Statement 2, 3 *etc*, Internet Engineering Task Force specifications covering the secure and reliable communication of business-to-business data via the internet using HTTP and digital signatures for document signing and encryption, with receipt confirmation returned after decryption and signature verification to guarantee successful delivery. Used for EDI message delivery, the content can be any standardised structured data (including ONIX messages). AS3 uses the FTP protocol instead of HTTP.

ASIN or Amazon Standard Identification Number: A proprietary identifier for products used internally within Amazon. The ASIN was originally the same digits as the ISBN-10, but this is no longer always true.

ASR or automatic stock replenishment: A tool used for demand forecasting, inventory management, and automated order generation to help keep stock levels from running low.

Attribute: In XML documents such as an ONIX message, text, numeric or other data contained *within* an opening markup tag (eg the *dateformat* attribute within <Date>). XML attributes usually carry information about how to *interpret* the data content of a data element. More generally, can be synonymous with ‘property’, ‘characteristic’, ‘data element’ or ‘data field’.

Audience Age Ranging: The grouping of ages into ranges (such as 0 – 2 years, 2 – 5 years etc), for the purposes of grouping books together based on their target audience.

Audio: Relating to sound.

Authority File: In library cataloguing and bibliographic data, a central list of, for example, contributor names. Used to ensure that contributors can be identified unambiguously and to highlight the single preferred form of a name that might have various forms or spellings. Any particular name may appear in different forms on different books, e.g. with or without Dr., with ü, ue or u, yet the shared contributor number from the authority file would make it clear that the names identify the same contributor. Authority files also help differentiate different contributors who share a name, and optionally can be used to resolve the real people behind pseudonyms. An authority file forms a type of controlled vocabulary.

Availability: Defines the status of a publication at a specific distributor or reseller. Availability should be updated as the underlying status changes.

B2B: An abbreviation for “business to business”. Describes commerce between businesses rather than between a business and an end consumer.

B2C: An abbreviation of business-to-consumer (B2C) which describes the practice of selling products and services directly between businesses and consumers.

Backlist: A published book that has been on sale for at least 6 - 12 months.

Barcode: Machine readable data printed as a series of black and white stripes on a product or on packaging. A conventional Bookland barcode on a book uses the EAN-13 barcode symbology and has the ISBN (in human readable form, with hyphens) above the stripes with the equivalent GTIN-13 printed at the bottom. The stripes represent the GTIN-13, not the ISBN, (though for modern books, the two are the same number). Some book barcodes – particularly in mass-market books – have a 5-digit price extension, extra stripes carrying the price or some publisher-specific code.

Best Practice: A set of principles that govern a concept, process or way of working and are recognised as delivering the best results. Best practice may become the standard or default over time.

BIC Subject Classification Scheme: An obsolete scheme. See Thema.

Book: A product in physical or digital form containing text, illustrations and/ or audio material.

Bookland: GTIN-13s are normally allocated nationally, with the first two or three digits indicating the country. Bookland is the fictional country to which the 978 and 979 prefixes used globally for ISBNs are notionally assigned. In this way, the range of ISBNs becomes a small subset of the larger GTIN numbering scheme.

Books Plus: Is a term that describes book products that contain more than just a book, for example a toy, a bracelet, a brush etc.

Buyer: The individual (end consumer) purchasing a book, book-related product or a package containing a combination of the two. Where a business is the “buyer”, this can relate to the purchasing of raw materials, finished products or services, and is delegated to an individual or department.

Catalogue: A publisher’s catalogue is the complete list of products it has available in print (i.e., frontlist and backlist items, but not out of print products). May also refer to a document listing selected products for marketing and promotional purposes.

Character Encoding: A defined list or repertoire of characters. A Character encoding then defines how this repertoire is represented by a computer. For example, ASCII lists a repertoire of 95 printable characters including space – plus a selection of non-printable ‘control characters’ including tab, new line, backspace/delete and so on – and encodes them using the numbers 0–127 (or 00000000 to

01111111 in binary, or 00 to 7f in hexadecimal). Latin-1 lists 191 characters and encodes them using the numbers 0–255 (00000000 to 11111111 in binary, 00 to ff in hex). Windows-1252 is a different encoding of around 215 characters also using the numbers 0–255 – and obviously this means that if some text is encoded using Windows-1252 and then displayed as if it were Latin-1, some characters will be displayed wrongly or not at all. See also Unicode, a character set of more than 140,000 characters.

Children’s: A means of grouping books together according to the intended audience or readership. “Children’s” books can be further divided according to age or subject genre.

Classification: The process of arranging books into one of a number of alike groups or classes, according to a particular attribute, and usually with each class in a particular classification scheme defined and arranged hierarchically. See for example “subject classification”, “harmonized system”. Categorisation is similar, though a single book may be a member of more than one category in a particular category scheme, and categories may not always be defined in a hierarchical fashion. In practice, the terms are often used interchangeably. Dewey is a subject *classification* scheme, since a book can have only a single Dewey number, whereas BISAC is a *category* scheme since a book can be assigned more than one BISAC subject code.

Code List: Term used in ONIX documentation for a controlled vocabulary or authority file. In addition, code lists define a language-independent notation – the code – for each term or concept in the vocabulary. Only the code appears in ONIX data. Some code lists have an implied hierarchy (for example list 150, where BA is clearly a broader term than BB or BC), making some lists taxonomies rather than simple vocabularies.

Collection: Fixed or indefinite number of products that share some collective identity such as a collective title. Members of the collection usually also have other attributes in common, such as product form or a branding or design style. A set or a series is a collection, but a collection could also comprise a less formal selection of products.

Collection Title: The title given to a collection of products. This can include a series title as well as the title of an editorial collection or some other form of collective grouping or branding.

Composite: Informal term used in ONIX documentation to refer to an XML markup structure that consists of an XML element or tag that contains only other data elements (and contains no data of its own). A composite acts as a wrapper around a set of closely related data elements, to indicate their semantic relationship and to enable them to be repeated in a neatly structured way. e.g. the repeatable <ProductIdentifier> composite contains three data elements, <ProductIDType>, <IDTypeName>, and <IDValue>.

Consumer: Someone who buys book or book-related products. Otherwise known as the “end-consumer”.

Contributor: Person or organisation – more generally, the party – responsible for creating the intellectual or artistic content of the product. ONIX is usually only concerned with contributors named on the product itself, and then only with their outward-facing public identity or persona. The contributor normally and automatically becomes the original copyright holder of the content (but not work for hire). The publisher acquires rights to exploit the intellectual or artistic content created by the contributor, in return for fees or a royalty.

Controlled Vocabulary: An exhaustive list of terms that can be used in a particular context or data element. Each term in the vocabulary should have an unambiguous definition or explanation and may include both preferred terms and less-preferred synonyms. Controlled vocabularies may be a ‘flat’ list of terms, or the terms may be arranged hierarchically – in which case the vocabulary is sometimes called a taxonomy.

Copyright: The exclusive legal rights to display, perform, reproduce ('the right of reproduction'), distribute and sell ('the right to distribute' and 'the right to make available'), modify or adapt ('the right of derivative works') or otherwise use original work or other intellectual property that is expressed in text, images, sound – a right enshrined in the Berne Convention, originally agreed in 1886 but revised and updated several times since – most recently by the Marrakesh Treaty.

The copyright in a work is held initially by the author or creator and can subsequently be passed on (e.g. to the author's estate), or licensed or assigned to publishers (and others) in a contract. Copyright in a new textual work usually persists for 70 years after the death of the original creator (occasionally slightly more) and proscribes exploitation of the work by those other than the copyright holder, licensees or assignees (collectively, rightsholders). In most jurisdictions, copyright (which is essentially a commercial right) is accompanied by inalienable Moral rights such as the right to be identified as the author, and protection for the integrity of the work.

Unlike rights over other intellectual property such as a patent or a trademark, copyright is automatic – you don't need to register your work to gain legal protection, though in some countries – particularly the USA – registration can be highly beneficial and in others, the moral rights must be explicitly asserted (for example, within the work itself). A few countries require works to be 'fixed' in a tangible form (e.g. published or made available in some format) before copyright applies. In others, fixation is not required.

The term of copyright has varied significantly across different countries during the last century, so 70 years after death is not always correct for older works. After expiry ('exhaustion') of the commercial rights, the work passes into the public domain. In some countries, Moral rights expire alongside the commercial copyright. In others, they are perpetual.

Copyright Exception: Certain groups, e.g. print-impaired readers, may hold a legal copyright exception and can make copies for personal use without obtaining permission from the rightsholders. Other uses of copyright material such as text and data mining (TDM) may also be allowed without explicit permission (e.g. for purposes of education, research, parody, for review and criticism, for digital backups etc) under 'fair use' or 'fair dealing' provisions of national law, but the scope and detail of these exceptions vary from country to country.

Core Product Information: The key characteristics that define a product.

Country of Manufacture: The country in which a product is physically manufactured (printed, bound, etc.), or very occasionally, the last country in which significant value was added (for example, if manufactured and packaged in different countries, it may be the country where the product is packaged) or where transformation took place (leading to the commodity code being changed). The Country of Manufacture is usually the country in which the HS Code last changed (e.g., when paper becomes printed books).

Country of Origin: A somewhat ambiguous term, frequently confused with or used as a synonym for "country of manufacture" or "country of publication". As a result, the term is avoided in ONIX. "Country of Origin" can refer to the "economic nationality" of imported goods according to the Rules of Origin applied by the importing country. Country of origin can also refer to the "country of exportation", the source of a particular international shipment of product(s), rather than where the products were published or originally manufactured. For example, for a physical book published in the United Kingdom but manufactured in Germany, the country of origin when a shipment of copies arrives in the UK could be the Netherlands – copies may well be transferred by road from Germany to Rotterdam before onward shipment to the UK. Equally, the country of origin could be Germany according to the UK's Rules of Origin.

Country of Publication: The country in which the publisher is based, and by extension, the country of

the national ISBN registration agency from which the ISBN should be obtained. Often also the country where legal deposit is mandated.

D2C or Direct to Consumer: The selling of goods direct to consumers by manufacturers or brands. Disintermediates intermediaries such as retailers and resellers.

Data Aggregator: An organisation that collects product metadata. This data is validated against recognised industry standards and made available on a commercial basis to other interested parties.

Data Element: In XML documents such as an ONIX message, a single XML tag and its content – text, numeric or other data contained between a pair of markup tags. Sometimes loosely termed a ‘data field’. *cf* attribute, composite.

Data Recipient: The recipient of product information. The recipient may be an intermediary or the ultimate beneficiary of the metadata. In either case, the recipient may accept the data as is, enhance the data or use it to supplement or update existing information.

Data Sender: As with the data recipient, the source of the product information can be the originator or an intermediary. Dependent on whether it is the former or the latter, the information may simply be geared to make the product discoverable or be enhanced to aid tradability.

Dependencies: Specific attributes of a product that are directly dependent on other attributes. For example, the availability status ‘in print’ is dependent on the publication date being in the past.

Digital: Recorded, stored, processed or reproduced electronically from numerical data.

Digital Asset Distributor or DAD: An organisation that facilitates the distribution of digital assets such as e-books to online retailers and libraries on behalf of publishers. The service may encompass a managed asset repository, file format conversion services, metadata and e-book distribution and aggregation of sales statistics. The publishers are not necessarily owned by (or own) the digital asset distributors.

Discount: In some countries, books have established wholesale or business-to-business prices. In others, business-to-business transactions are based on a discount from the fixed or recommended retail price (business-to-consumer price) agreed by the parties. The trade discount given by a distributor or wholesaler varies from retailer to retailer (bigger retailers sometimes get more discount) and from book to book (discounts are often greater on mass market fiction than on specialist non-fiction). Assuming the books are sold to end customers at their normal retail price, the trade discount represents the retailer’s gross margin (sales revenue minus cost of goods sold). More loosely, Discount can also refer to books sold at retail for less than their recommended retail price (which reduces the retailer’s margin).

Discount Code: Index into a table of discounts shared in advance between supplier and retailer. Since each retailer may have a unique table, the actual discount percentage can be communicated easily without revealing commercially sensitive information.

Distribution Centre or DC: A publisher’s, distributor’s or wholesaler’s warehouse.

Distributor: Organisation that holds the primary stock of books and is responsible for fulfilment (of trade orders, or of orders from a specific type of customer, e.g. education) in a particular territory or market on behalf of the publisher. Wholesalers and retailers may act as intermediaries between the distributor and the end customer. Many large publishers own or operate their own distributor and hold stocks themselves. Other publishers (a distributor’s ‘client publishers’) appoint a single, exclusive distributor per market or territory (and this exclusive distributor is sometimes termed the Vendor of

record for that market). Some publishers prefer to appoint multiple non-exclusive distributors. In North America in particular, distributors may also have a role in marketing, promotion and sales of the book to wholesalers and retailers – that is they act as a sales agent – and in ONIX these distributors should be named both as <PublisherRepresentative> and <Supplier> in the relevant market. More generally, distributors frequently take on this broader role outside the home market of the publisher. In France, this distinction is expressed in the terms « diffuseur » which takes a role in marketing and sales, and « distributeur » which concentrates solely on fulfilment.

Dramatised: In audiobooks, a book or other text that has been adapted to be presented as a play, rather than simply read aloud, and usually featuring a cast of audio actors rather than a single reader.

EAA or European Accessibility Act: An EU Directive which must be enacted into EU member states' national law by mid-2022 and take effect in 2025. It mandates a uniform high level of accessibility for goods and services across the EU, including *inter alia*, e-books and other digital products.

EAN: Former European Article Number, later International Article Number (though still abbreviated to “EAN”).

E-book: An electronic book. Has the attributes of a physical book (cover, table of contents, chapters with text and images, product identifier), but is in digital rather than physical form. E-books are available in a number of different formats (including EPUB, Kindle and PDF) and can be downloaded to smart phones, e-readers, tablets, computers or accessed online.

EDI or Electronic Data Interchange: A structured, technical format for the exchange of time, business or financially critical transactions in an automated and standardised form between computers. Transactions are processed in batches rather than in real time.

EDIFACT or Electronic Data Interchange for Administration, Commerce and Transport: The UN standard for electronic data interchange of time, business or financially critical transactions in an automated and standardised form between computers.

Edition: Generally, means all copies of a book that contain the same content, usually published by the same publisher. More loosely, an “edition” may be a product produced for a specific sales channel or market segment – a book club edition, library edition, large print edition and so on. Different editions may be distinguished by their content (by the addition, revision or removal of material), or more occasionally by some other aspect of their product form or nature.

End-Consumer: Individual purchaser of book or book-related products. Otherwise known as the “consumer”.

File: An electronic container storing information.

Fixed Retail Price or FRP or Fixed Book Price or FBP: In certain countries, for example France or Germany, there are legal restrictions that prevent retailers selling books significantly below the list price set by the publisher (or occasionally the importer). This usually applies for a year or two following publication.

Foreign Currency: The money used in another country for the buying or selling of goods.

Foreign Currency Exchange Rate: The cost to buy the currency of one country using the currency of another.

Format: Describes the form the product takes.

Freight Forwarder: An organisation specialising in domestic and international transport and logistics, arranging transport of cargoes of products from the manufacturer to a distributor, wholesaler or retailer along the supply chain, on behalf of its client publishers or manufacturers (the shippers or consignors, who ultimately bear the costs of transport). A freight forwarder does not itself move goods, but contracts with carriers to do so, and is usually responsible for arranging Customs clearance and other administration of the shipments. Freight forwarders may also arrange consolidation of individual small shipments into larger cargoes to reduce overall transport costs.

Frontlist: The range of forthcoming and recently published titles that publishers are actively selling and promoting to the book trade. Soon after publication they will become publishers' "backlist" (q.v.) titles while newer books replace them.

FTP or File Transfer Protocol: A standard method of transferring files across the internet. cf HTTP.

FTPS: Version of FTP in which the data transferred is securely encrypted while in transit across the network. FTPS (more properly 'FTP over TLS') uses a digital certificate for network authentication and for establishing the encrypted link to maintain the privacy and trust in the integrity of the data. SFTP (more properly 'SSH-FTP') is a different approach to the same requirement, and SFTP and FTPS each has its own benefits. See also HTTP and HTTPS, which can also be used for file transfers. See also AS2, AS3.

Fulfilment Centre: Warehouse location where the product is packaged and dispatched from.

GLN or Global Location Number: An international standard identifier for a physical trading location or (loosely) for an organisation at that location. Well-established within e-commerce and physical logistics, and not in any way specific to the publishing industry. cf SAN. Although the GLN system is administered by GS1, there is only limited central coordination of GLN assignment, and a single location may have more than one GLN. Note that GLNs are 13-digit numbers and must be distinguished from GTIN-13s by context.

GS1UK: A not-for-profit standards organisation. BIC is licensed by GS1UK to allocate SANs and GLNs.

GTIN or Global Trade Item Number: A numbering scheme for tradeable items and consumer products of all types in the supply chain. The GTIN identifier scheme is administered by GS1. GTIN-13s, formerly known as "EANs" (European Article Numbers) are used globally to identify a wide range of retail items. The barcode symbology used to represent GTIN-13s is still referred to as 'EAN-13'. Thirteen-digit ISBNs are a small subset of the GTIN-13 number scheme.

Harmonized System or HS: A set of codes defined by the "World Customs Organization" for commercially traded products. These 6-digit codes are the basis for the commodity codes on the type of product and component materials. Pulp, paper, books and periodicals are all found in Chapter 49 of the Harmonized System, and most ordinary books are classified as 4901.99. The HS is used as the basis for more elaborate national schemes such as the UK's Trade Tariff or the EU's TARIC codes, and the schemes are used to administer Customs tariffs, duties and quotas for import and export, and for the compilation of trade statistics. They can also be used to administer internal tax rates.

HTML or Hypertext Markup Language: The markup system used for simple web pages. Sometimes refers specifically to HTML version 4, standardised by the W3C in 1997, but in other contexts encompasses HTML5 too. It uses simple XML-like tags to add structure to plain text, for example by surrounding third-level headings with `<h3>` tags and by marking paragraphs with `<p>` tags. But although XML-like, it does not fully conform to XML syntax, as certain HTML end tags are optional (e.g. `</p>`, `` or the `/` character in `
`) and tags may be upper or lower case.

HTTP or HyperText Transfer Protocol: A standard method used for transferring files or data across

the internet. HTTP is used to transfer normal web pages (at their most basic, these are just files that use HTML markup) from web server to browser – but HTTP can be used to transfer other types of information too. cf [FTP](#), [HTTPS](#).

HTTPS: A version of [HTTP](#) in which the data transferred is securely encrypted while in transit across the network. It uses a digital certificate for network authentication and for establishing the encrypted link to maintain the privacy and trust in the integrity of the data.

Identifier: Effectively, a persistent ‘name’ or ‘label’ for some entity like a product, a work, a location – or indeed a name – where the label is unique within a given context. Identifiers are often (but not always) in a tightly controlled alphanumeric format, and sometimes contain a check digit to help error detection. Standardised identifiers (for example the [ISBN](#) or [ISNI](#)) generally provide global uniqueness, there is often a minimum set of [metadata](#) associated with the identifier, and the identifier and metadata are sometimes managed in a centralised registry. Other identifiers use decentralised registries. A well-managed registry engenders trust in the identifier, its persistence, and the likely future sustainability of the identifier scheme itself.

Although many identifiers are constructed using a ‘recipe’ (something like ‘four digits for the year, three for the [publisher](#), then five more digits and a check digit’), it is best to treat them as dumb labels without internal meaning, intelligence or affordance.

For standardised identifiers such as the ISBN or ISNI, the nature or scope of the identified entity is well defined and understood, so they enable unambiguous communication between organisations within the supply chain.

Impression: Refers to a print run of a specific number of copies.

Imprint: A brand within a publisher. The [publisher](#) is the legal entity, whereas the imprint is not. An imprint will specialise in a specific range of books, e.g., Classics, [Children’s](#), Sci-Fi or Lifestyle. Often (but not always) imprints share the same editorial and marketing teams. Often hardback and paperback editions of a book will be published under different imprints yet remain owned by the same publisher.

In Print: Active, available, commonly abbreviated to IP. Status implying the product has been published and is orderable from the [publisher](#) or publisher’s primary [distributor](#). Note that this does not imply the product is *immediately* available – it may be temporarily unavailable. Conversely, [out of print](#) does not mean the product is necessarily unavailable – there may still be stock available within the [supply chain](#). Out of print means only that the publisher or the publisher’s distributor will no longer accept further orders for the [product](#).

Occasionally used to indicate numbers of copies sold (‘4 billion copies in print’).

ISBN or International Standard Book Number: A thirteen-digit product [identifier](#) for books in physical, [e-book](#) and audiobook formats. Administered by the International ISBN Agency and numerous affiliated national agencies (including Nielsen BookData in the UK and Ireland). Publishers apply for ISBN allocations through their local ISBN Agency. Used for book discovery, ordering and tracking sales, the ISBN is unique to a specific title, edition and format combination, and cannot be re-used.

ISNI or International Standard Name Identifier: A standard identifier for public identities or personas of parties (people, organisations) involved in creative activities. It can also be applied to brand names (imprints) and to fictional characters. An ISNI can provide an unambiguous way of identifying contributors, imprints and publishing companies. An ISNI consists of 15 decimal digits plus a final check digit which may be 0–9 or X. ISNI is an [ISO](#) standard, and the ISNI system is managed on behalf of the ISO by the ISNI International Agency (ISNI-IA) registration authority and a range of

ISNI registration agencies.

ISO or International Organization for Standardization; The non-governmental global standards setting body of which many national standards bodies are members. ISO Technical Committee 46 Subcommittee 9 is ultimately responsible for many standardised publishing identifiers including the ISBN and ISNI, plus the foundational *Principles of Identification* standard, but ISO sets standards for many other aspects of the publishing industry too.

Keyword: Word or phrase chosen to describe or associate with the content or theme of a book. Keywords do *not* conform to a controlled vocabulary, but may be any natural language word or phrase, such as names of characters in a novel, locations, narrative themes, terms of art etc – any relevant word that is likely to be the target of a search.

Language: The language the product is displayed in.

Legal Deposit: Legal requirement and administrative process whereby publishers lodge a copy – sometimes multiple copies – of every publication with a national library or with other repositories.

Levelling: Measurement or assessment of the complexity of text and its context, or the reading ability required for comprehension. Also, the arrangement of multiple books (e.g. in a collection of educational readers) into a sequence of progressively increasing complexity.

Long Term Archive: A service that safeguards access to e-journals, e-books and digital collections to ensure that the content remains accessible to and usable by researchers, scholars and students in the future.

Manufacture: The production or creation of a product, including the printing and binding of physical books.

Manufacturer: A business producing goods for sale. Includes printing and binding.

Markup: Labels, delimiters or tags within a document that define its structure or meaning. In XML and HTML, markup tags are placed between < and > symbols. Tags are often paired to indicate the beginning and end of a particular data element within the document, so top-level heading text in HTML is contained between <H1> and </H1> tags. Markup can be described as semantic or presentational, but in practice is usually a mixture of both.

Market: A geographical area within which commercial arrangements for distributing and selling a product are consistent – usually with a single exclusive distributor and single availability date across the market.

Metadata: Strictly, data about other data. More usefully in the context of the book and e-book supply chain, metadata can be thought of as all the data used to describe and trade products through the supply chain. This encompasses both simple, structured and factual information like titles, author names, distribution arrangements and prices, and richer, more complex descriptive data, classifications of various types and even parts of the book itself (a table of contents can be seen also as valuable descriptive metadata). ONIX messages are a method of communicating standardised and highly structured metadata from one party to another within the supply chain. Some organisations might also consider internal workflow information to be part of the product's metadata.

National Library: The library established by a government as a country's preeminent repository of information.

ONIX or ONline Information eXchange: The international and most widely used standard for capturing

and communicating book product information electronically. This includes both physical and digital products.

ONIX Feed: Arrangement and supply of an ongoing series of ONIX messages exchanged between data sender and recipient, generally on a daily or weekly schedule or 'as needed'. Once a data feed has been established with an initial message containing the full set of product records, subsequent messages in the feed normally contain only product records for new products and updates (more accurately, replacement data) for existing products.

ONIX Message: A complete ONIX data file, generally one in a series of messages (an ONIX feed) passed between a data provider and a data recipient. A single message may contain one or many product records, or none.

Open Access or OA: Licensed on 'open' terms following the recommendations of the Budapest Open Access Initiative and later policies, for example under a Creative Commons license, which typically mean that a work or published product can be read, used and re-distributed freely – free of charge, and free of at least some of the usual copyright restrictions. The particular license chosen may still require attribution, prevent the distribution of derivative works or commercial exploitation, or impose other restrictions.

Out of Print: A publisher may declare a product out of print (or OP) to indicate it (or its primary distributor) will no longer accept orders. This usually also means copies sold to retailers on a sale or return basis are no longer returnable (or may be returnable only for a short period after the product is declared OP). However, out of print does *not* mean 'unavailable', as there may still be many copies remaining for sale in the supply chain. Out of print can also be applied to a work rather than an individual product, which implies *all* manifestations of a work are out of print. Out of print is generally viewed as a 'final' status and a particular product (a particular ISBN) cannot revert to an active status – though a work considered out of print can revert if a new manifestation (with a new ISBN) is made available. Publishers *do* occasionally revert products from OP to a 'forthcoming' or active status (without a new ISBN), and even set a new 'publication date', but this is seen as exceptional and is not recommended practice.

Overstocks: Books that are printed but not distributed or are distributed but not sold. Overstocks can occur at various points in the supply chain (e.g., at the distribution centre or the bookseller). They are often sold at a lower cost or returned and/or destroyed depending on the specific situation.

Packager: An agency contracted by the publisher to produce a book, usually including text creation, editing, design and illustration, but not the manufacturing of the final product.

Packaging: The combining of a book and companion book-related product into a single package for sale.

Page Extent: Page count, the number of pages in a book (or occasionally, the number of words or the running time). There are different methods of measuring the extent.

Physical: As opposed to digital. Describes a product that is tangible.

Platform: A digital service making products available to consumers.

Post Publication: After publication date.

Pre-Publication: In advance of publication date. Often abbreviated to 'pre-pub'.

Price: The amount a publisher, supplier or retailer chooses to sell a product for. A product's 'price' can

refer to a recommended retail price, actual selling price or promotional price.

Printer: Reproduces text and images in books and periodicals.

Printing: The physical reproduction of text and images in books and periodicals.

Print on Demand or POD: The manufacture of a book – often using xerographic (dry, toner-based) or inkjet printing – in response to a customer order. POD copies may be drop-shipped to the retail customer or fulfilled via the retailer.

Print Run: The number of copies printed in a single impression. Historically, this was an edition, and this sense is still used in book collecting (i.e., a ‘first edition’ is a copy taken from the first print run).

Product: Any separately tradable book-form item. This includes physical books, e-books and audiobooks. In the book trade, each separate product is normally identified by a different ISBN to facilitate discovery, ordering and sales analysis.

Product Record: The complete collection of metadata relating to a single product, provided within an ONIX message. A Product record is contained between <Product> and </Product> XML tags.

Production: The act of creating or manufacturing a product from raw materials. Also involves the application of industry expertise.

Publication: Another term for a book or other product designed to be read.

Publication Date: The date on which a book is officially published. For digital audiobooks, this is usually (but not necessarily) the date on which retail consumers can first download or stream the entire audiobook. See also Sales Embargo Date.

Publisher: A business responsible for bringing books to market. Publishers will usually have their own editors, marketing and publicity teams as well as a unique stable of authors. Publishers are not always responsible for managing distribution or returns.

Reading Age: Measure of a child’s reading proficiency or the proficiency required to read and comprehend a text, expressed as the age of a child of average reading ability.

Related Product: The term has two distinct meanings. In an “ONIX” composite context, it is used to specify the ISBNs of any similar or related products to that in the main product record. More generally, the term describes non-book items made available for sale in the book industry because they are considered to be of relevance to consumers or likely to trigger impulse buys.

Retailer: The organisation selling goods to the end consumer. A retailer can sell through physical (“bricks and mortar”) bookshops, an online store or may operate a direct sales operation (such as a book club).

Rights: A general term covering copyright, moral rights and other intellectual property rights, plus contractual rights such as the right to distribute or sell products. So called “volume rights” give the publisher the right to publish and sell products based on a copyright work and are sometimes divided by language and geographical territory. Subsidiary rights – initially attached to the volume rights but often sub-licensed by the volume rights holder to another publisher – sometimes include the right to publish specialised manifestations such as audio.

Royalty: A percentage of the recommended published price of the book or a percentage of the net receipts payable to the author, calculated on each copy sold, or in relation to income received from the

licensing of rights of the book.

RRP or Recommended Retail Price: Often just RRP, or occasionally termed a Suggested retail price, or SRP. Price chosen and recommended by the publisher for sales to the consumer. The retailer does not *have* to use this price and may choose to sell the product for a lower (or higher) price – the Actual selling price or ASP. But royalties paid to the contributors of the book are often based on a percentage of the RRP, even though the Actual selling price is different. In countries where RRP's are the norm, the Agency model may still allow publishers to exert direct control over consumer prices. In some countries, retail prices set by publishers are *fixed*: by law, retailers may not reduce (or increase) the Fixed retail price or may do so only within a fairly narrow band or only after a certain time has elapsed since publication.

Sales Department: The team within an organisation responsible for marketing, promotion and financial deals involving specific goods or services.

Sales Embargo Date: The earliest date on which retailers may sell a book product to consumers. Distinct in meaning from the publication date, though it is usually the same actual date. This is the date which determines the release of the book by retailers to consumers. All books have a publication date, but only some have a sales embargo date.

Sales Rights: The commercial rights derived from the publisher's publishing rights that a publisher confers on its distributors, wholesalers and retailers, allowing them to trade in and make the product available to customers. Note the contrast between the publishing rights and the sales rights: publishing rights concern where (and perhaps *when*) a publisher has the *right* to publish and sell a product. The sales rights are where the publisher *chooses* to make the product available (*i.e.* chooses to *exercise* those publishing rights). Clearly, the sales rights must be a subset of (or the same as) the publishing rights.

Underlying these publishing rights and sales rights are more specific rights, ultimately derived from the creator's copyright. In particular, the publishing rights indicate that the publisher holds the 'right to reproduce', the 'right to distribute' and the 'right to make available', and the sales rights indicate the publisher is willing to pass (some of) these rights on to its supply chain partners. Without holding and passing on these necessary rights, the publisher cannot legitimately express the sales rights.

In ONIX, only the *sales rights* are described in detail, and the publishing rights themselves are not made explicit (at least in part because the publishing rights usually pertain to the work, not the product). In ONIX, the <SalesRights> composites list the set of countries and regions where the publisher is exercising its rights to make a product available. A <ProductSupply> composite and its enclosed <Market> composites can detail the subset of countries and regions where those rights are conferred upon a *particular* set of distributors, wholesalers and retailers. Another market might have a different subset of the sales rights conferred upon a different group of suppliers. When broken down by market, these are often termed distribution rights.

Sales Tax: A tax levied as a percentage of retail sales to the consumer or end user, cf. VAT, which is levied incrementally at all points in the supply chain. Sales taxes are levied by most US states and some Canadian provinces. Rates vary, and additional sales taxes may be levied by city and local governments. Since the total retail price – inclusive of tax – varies according to the exact location of the retail sale, advertised prices in those countries do not include the sales tax element, and tax is added at the checkout.

SAN or Standard Address Number: National standard identifier for a trading location within the supply chain. SANs are used in the United Kingdom, where the registry is maintained by Nielsen BookData. In contrast to the GLN, the SAN is unique to the publishing industry, but is well established in book-related e-commerce in North America and parts of Europe to identify distribution locations,

customer delivery addresses *etc.*

Schema: An XML schema formally defines the set of markup tags that may be used in a particular type of XML document, whether each tag is mandatory or optional, and their order and nesting. But *unlike* a Document Type Definition, a schema also constrains the data types and values that may be used within the data elements in the document. It can require that a particular tag contains an integer, or a date, or set a limit on the length of text. And an XML schema can define lists of allowed values ('enumerations', controlled vocabularies, or in ONIX terminology, code lists) that can be used in a particular data element. Validation is the process of checking a particular XML file meets the requirements defined by a particular schema.

The normal or 'classic' ONIX XSD is based around XSD 1.0. A 'strict' XSD based on XSD 1.1 technology is also available, and this checks a further range of data types, business rules and other requirements, although it is not compatible with all XML validation scenarios.

Series: A sequence of books that have shared characteristics and are identified collectively as a group.

Set: Finite number of products published simultaneously or over a definite period of time, with a shared identity such as a 'set title'. The products are usually of similar product form and share a distinctive branding or design style. The products in the set may be available individually, or the set may be a single product, or both. In ONIX, a set is a type of collection. *cf* series.

SFTP: See FTPS.

Short-Run Printing: Production at the behest of the publisher (not end consumer) that refers to quantities generally considered not economical for litho/offset printing. Usually employs toner or inkjet technology similar to that of POD. These copies are then warehoused and distributed in a conventional manner (though warehousing for such small numbers of copies may be at the printer, rather than a dedicated distributor's or wholesaler's warehouse).

SLA or Service Level Agreement: The agreed quality of service (often quoted in terms of time to react, time to complete a task, acceptable technical standards *etc*) to be provided by a service department or an external partner.

Standard Operating Procedure or SOP: A set of written instructions required to carry out a routine procedure in an effective and efficient way.

Subject Classification: Systematic arrangement of books according to what they are *about*, according to some subject scheme. Many schemes – for example Thema, BISAC, Dewey – list all potential subjects or topics hierarchically, and each subject is associated with a code (or controlled vocabulary). Most strictly, classification implies a book may be a member of only a single subject group (*i.e.* the groups are mutually exclusive), whereas categorisation implies a book may be assigned to many categories. In practice, however, 'categorisation' and 'classification' are often used interchangeably. Dewey is a subject *classification* scheme, since a book can have only a single Dewey number, whereas Thema and BISAC are *category* schemes since a book can be assigned more than one Thema or BISAC subject code.

Supplier: The source of a product or service, usually on a commercial basis.

Supply Chain: In the book industry, this comprises all the organisations, individuals and activities that are involved in the creation, manufacture, distribution, sale and return of books and book-related products. Also includes the support functions such as Marketing and Finance. The supply chain model will vary according to the type of product (physical/ digital) and who the consumer is (such as B2B,

B2C or B2G).

Sustainability: Indicates the ability or goal to maintain or improve upon something desirable or beneficial over a long period of time. This is usually considered from environmental, social and economic viewpoints. It can encompass the avoidance of the depletion of natural resources or the degradation of the natural environment, development towards a more equitable society that provides for and protects vulnerable communities or building successful business models to secure the long-term future of organisations and business systems.

Tag: Can refer to either the markup elements of HTML, XML or ONIX (e.g. ‘the <ProductForm> tag’), or to keywords associated with some content that are used to classify the content (e.g. ‘blog posts tagged with “ONIX”, or ‘a tag cloud’). It is common to confuse these two very distinct meanings, particularly as classification tags can sometimes be embedded within markup tags.

Tax: A government-imposed levy on business, goods, services or transactions.

Taxonomy: A classification scheme, where controlled vocabulary terms or concepts are arranged in a hierarchy of classes and sub-classes. Strictly, a particular entity can only be attached to a single class or term within the taxonomy, though this is not always rigorously applied (for example with many ‘subject classification’ schemes – these are really subject *category* schemes, where entities can be assigned to multiple categories). Where the vocabulary terms are related not just hierarchically (i.e. broader and narrower terms) but also by *non-hierarchical* links of association (‘related to’) and equivalence (‘same as’), and terms are accompanied by a richer range of usage notes, the scheme is often called a thesaurus. A formal representation of the concepts and the relationships is an ontology.

Territorial Rights: Describes where a product may or may not be sold (or be available to borrow).

Thema: An international subject classification scheme for the book trade. Used to improve merchandising, discovery, sales and market intelligence.

Third Party: An external party contracted to provide a service or range of services requiring specific technical, standards and best practice expertise to a business.

Title: The name of a book, as listed on the title page or cover, or one of the variants of that (former title, abbreviated title, alternative title *etc*). This might also include any collection title. Titles often incorporate a volume or part number when a work is split into separately bound books, a year or year range for annually issued publications, or a subtitle which may amplify or clarify the title. The title does *not* include any promotional cover lines, endorsements or marketing messages that might also appear on the book cover. More loosely, can imply a *group* of products that share the same name and contents – more akin to a work.

TRADACOMS. Trading Data Communications Standard. A UK EDI standard that is still commonly used, though now superseded by “EDIFACT”. Used for the exchange of time, business or financially critical transactions in an automated and standardised form between computers.

Trade: The book trade – the business and commerce of book publishing, manufacture, distribution and selling, in phrases like ‘trade-only’, ‘trade-pack’, ‘trade association’, trade discount *etc*. A trade book – fiction and general interest non-fiction books primarily for adult consumers, sold through ordinary retail bookshops.

Unicode: Comprehensive set of characters standardised as ISO 10646, and hugely improved over ASCII and Latin-1 to Latin-10, which are focused on basic English text and a broad range of European scripts respectively. The Unicode character set, or UCS, includes the characters required for all common written languages and writing systems, including non-European scripts like Arabic,

Devanagari or Chinese, ancient and historical scripts, and a vast range of mathematical and other symbols – even emoji. However, *displaying* these characters is font-dependent too – the data may be in Unicode, but if your font does not contain the right characters, you still can't see it!

Each character in Unicode has a 32-bit number or codepoint (*cf* ASCII, which uses 7 bits (binary digits) for each codepoint, or Latin-1, 8 bits per codepoint, and are therefore limited a much smaller range of characters), so in principle Unicode files are four times the size of Ascii or Latin-1 files for the same text content. However, UTF-8 is a special encoded form of Unicode that reduces the file size to something more manageable.

Validation: The process of checking an XML document (such as an ONIX file) first meets the requirements of the XML syntax (a single root element, all tags explicitly closed *etc*), and second meets the requirements of a particular schema – for example the ONIX for Books XSD, meaning that it's 'valid' according to the structure or grammar of a particular schema. Special XML parser software is required to validate an XML document. Note also that validation is not necessarily enough – there may also be various restrictions and business rules on validity that may not be possible to check via a schema.

VAT or Value Added Tax: Also known in some countries as “Goods and Services Tax” or “GST”. It is levied as a percentage of sales (both B2B and B2C sales) and paid by companies to the tax authority based on the difference between the tax value of their sales and the tax value of their purchases. This is collected incrementally from all parts of the supply chain. VAT/ GST rates vary from country to country. In some countries books and other cultural goods attract a special low rate. In the UK, VAT on physical and digital books (but not audio) is currently levied at 0%.

Vendor: An organisation offering a product or service for sale or loan on a commercial basis. The vendor fulfils roles in different supply chain models. These include direct to consumer, direct to government and business to business.

Warehouse: A building designed to store products prior to sale and distribution.

Wholesaler: In the book industry, a business that has the attributes of a distributor and a retailer, and often acts as an intermediary between the two. It sources stock for quick order fulfilment, not just large quantities of the faster selling titles, but also slower selling, specialist titles.

Work: Distinct intellectual or artistic creation, from which several products (or manifestations in <indecs> terms) may be created; roughly synonymous with title, where hardback, paperback and e-book manifestations of that title are all versions of the same work. Textual works may have a work identifier of some type. Works are the subject of intellectual property rights. Modifications to a work, such as translation, abridgement or significant revision (e.g. to create a new edition) or addition of further material, create a new Derived work, and thus multiple works can be arranged in chains or networks of derivation. As such, ONIX uses the <indecs> view of a work.

XML or Extensible Markup Language: A set of codes or tags that describes text in a digital document. The structured human/ computer readable format is used in the book industry for the latest standards such as ONIX, EPUB, EDItX and some application programming interfaces.

XSD or W3C XML Schema Definition: The language for defining XML schemas.