

The identification of digital book content

A discussion paper commissioned by the Book Industry Study Group and Book Industry Communication and prepared by Michael Holdsworth

Executive summary

Although there are strong similarities between the identification needs of physical books and of digital book content in the supply chain, new business models and new delivery channels challenge existing practice. There is a pressing need for clarity on the use of standards for the identification and description of digital products.

The International Standard Book Number (ISBN) is considered fit for purpose for trading, discovery and reporting of digital products within the supply chain.

Every digital manifestation traded separately should be assigned an individual ISBN and publishers should adopt the principle that products should be separately identified to the extent that they need to be so identified.

The Digital Object Identifier (DOI), though not in any way a substitute for an ISBN, is a valuable additional identifier which enables content to be “discovered” on the Web.

Publishers and others should explore the possible opportunities for using the International Standard Text Code (ISTC) as a means of collocating different manifestations of the same textual work.

Publishers should review their practices relating to the assignment of identifiers, particularly where a digital product is traded through an intermediary and accurate reporting of sales is required for royalty payments or management information.

The practice of using a single ISBN for all digital manifestations of the same work is strongly discouraged.

The use of “ISBN-like” 13 digit identifiers other than those properly assigned by the ISBN agencies is strongly discouraged in all circumstances.

Introduction

This paper has been commissioned by the Book Industry Study Group (BISG) in the USA and Book Industry Communication (BIC) in the UK. It has two purposes. First, it is intended to stimulate discussion about how digital book content should be identified. Second, it is intended to encourage further work on the development and implementation of identification standards for such content, particularly (but not exclusively) in the books supply chain.

Up to now, the sale of e-books and digital content has achieved a degree of scale only in the academic and professional sectors. One part of those sectors – scholarly journals – has moved during the last fifteen years to a state of near-maturity, where all day-to-day access and usage is online, with print relegated mainly to archival uses. With the coming to market of new reading devices, and the likely emergence of a completely new supply chain for digital delivery of everything from complete e-books to individual chapters or smaller fragments, it appears that the revolution in electronic publishing is set to extend through all sectors of the book industry.

The challenges posed by digital content

For book publishers taking their first steps in e-books, the commercial mechanics of a simple product sale to a consumer of an electronic version – whether downloaded to a PC or a handheld device, or made available online on a perpetual basis – are broadly similar to those of selling a printed book. However, once book content is in digital form, an almost infinite number of new business models become available, both for online and for offline consumption.

Many of these new models have been tried and tested within the academic and professional sectors. Such models range from the outright perpetual purchase of a digital e-book version to time-delimited rental or subscription and pay-per-view. Within those models there can be an almost infinite range of usage rights conferred or denied to the user. This is the controversial universe of publisher-controlled restrictive Digital Rights Management. DRM might constrain or allow the right to move the content between devices (from a PC to a handheld; or among multiple devices simultaneously); or the right to print pages or whole copies; or the right to capture content to the clipboard via copy-and-paste. And within those usage rights there can be an almost infinite range of controls: allowing print or copy for a set number of pages only, or a set percentage of the whole, or a maximum number of pages at one time, or a maximum number within a week or a month or ever. The combinations are almost limitless and all of them pose issues of identification both for trading in the supply chain and for the maintenance of systems and records by publishers.

With new business models comes the need for new specialist services from vendors. Sophisticated systems for managing workflow and content (Digital Asset Management systems – DAMs) are being developed and deployed, as are systems for aggregation, monetization and distribution (Digital Asset Distributors – DADs). Some of these systems are being developed by new industry entrants, and they and the traditional publishers and booksellers they will serve are beginning to demand a standard framework for the identification of digital assets.

The need is urgent. In the business-to-business environment, booksellers, libraries and publishers need to know which versions are available and which they are promoting, ordering, trading and buying. Consumers need to know which digital versions of titles are available, and whether these are compatible with their needs or with their software or hardware devices, and they will need to understand what their usage rights will be. There is therefore a pressing need for clarity on the use of standards for the identification and description of digital content as this new market evolves.

The existing identifiers

The International Standard Book Number (ISBN)

The international book industry, since it became computerized in the 1970s, has managed almost all its identification needs around the International Standard Book Number (ISBN). The ISBN has been a remarkable success story, as is clear from the fact that 166 countries are now members of the system. The international success of the ISBN and the proliferation of new e-publications requiring identification contributed to the decision in 2005 to replace its ten-digit form with the EAN/ISBN-13. This solution, implemented in January 2007, has removed the fundamental concern about simply running out of numbers, though some deep anxieties remain about numerical bloat within publisher, distributor and bookseller systems. And self-evidently, the issue of managing and identifying many hundreds of thousands of new digital products alongside of print will be the same challenge for the industry, regardless of which identifiers are employed.

During the process of revising the ISBN standard in 2005 the opportunity was taken to examine and redefine the scope of material eligible to be identified by an ISBN. The new standard for the first time explicitly incorporated electronic publications, chapters and parts of publications, and journal articles (when purchasable separately).

The ISBN is a supply chain identifier. It exists to identify each separately tradable “manifestation” of a publication. Its main purposes are to facilitate the trading of books, enable and support the discovery of books in bibliographic databases and library catalogs,

support the use of retail point-of-sale systems and the compilation of cross-industry sales data, and support rights management and national lending rights schemes. It was not designed to identify contracts, royalties or the parts of an individual product in a data repository, though some publishers do indeed use the ISBN for these functions. In fact, in many publisher systems the ISBN has become the de facto key product attribute, tying together sales, distribution, physical logistics, production workflow management, cost allocation, contracts management, rights sales, royalty payments and most elements of financial analysis. The success of the ISBN over more than thirty years has led to it being used in ways for which it was not designed or authorized.

The key question, of course, is whether the ISBN is the best possible identifier for digital publications in all their various forms. It clearly has one important deficiency: its inability to act as a tool for collocating or linking together different manifestations of the same content, a function that is becoming critically important as multiple printed and digital versions of the same title proliferate in systems across the book industry.

Publishers, booksellers, librarians and bibliographic agencies have traditionally all had their own makeshift solutions to this challenge – how to link together all the ways a customer might be able to buy specific content: in hardback, in paperback, in large-print, as an audio book, and possibly across translated versions in different languages. Various mechanisms – title-relationships, associated titles, the ONIX data structures – are currently employed, but more often than not these are manual or partly automated rather than fully computer-enabled.

The ISBN's deficiency as a tool for collocation matters less in view of the fact that a new international standard – the International Standard Text Code (ISTC) – has been developed and is expected to be launched in the middle of 2008. The pressing need for a standard tool for linking different printed and digital manifestations to a common title makes the ISTC potentially the most important standard for our industry since the ISBN.

Is the ISBN to be recommended as the identifier the book trade should be using for the trading of digital product? The answer must, despite some minor deficiencies mentioned above, be an unequivocal yes. In fact, it fulfils almost all of the needs of digital supply very effectively. It ties together digital product with physical product and there is no shortage of numbers to cope with any proliferation of versions and fragments. In most situations, there is no difference between using separate ISBNs for hardback, paperback, audio book, large-print, new edition and so on in the physical supply chain and using a separate ISBN for every digital manifestation which is traded. There are some cases, certainly, where new issues arise in the digital supply chain, and these will be considered later in this document. However, for almost all practical purposes, publishers are advised to assign ISBNs to their tradable digital products exactly as they assign

them now for printed books, and publishers are advised not to seek new proprietary solutions, which eventually could destabilize and undermine the supply chain.

Publishers – as the originators of product within the supply chain – are responsible for deciding which of their products will be assigned ISBNs, and we recommend that they assign ISBNs when the following requirements apply:

- The requirement for the product to be directly traded
- The requirement for the product to be “discovered”
- The requirement for transactions to be reported to them

We recommend publishers adopt the principle that products should be separately identified to the extent that they need to be separately identified. In other words, separate identification of a product is necessary only when it is important for trading, accounting or recording that it should be identified individually.

The Digital Object Identifier (DOI)

The Digital Object Identifier (DOI) has often been described as the ISBN of digital content. This is misleading and unhelpful. The DOI was never intended to function as a supply chain identifier for tradable content. Its value is that it can be used to point to any resource, physical or digital, by providing a publisher-controlled and moderated, fully dynamic, and updatable “destination” on the Internet where fuller descriptive information, additional or related content or ecommerce options can be provided.

Book publishers may increasingly recognize the benefits of assigning a DOI – as an aid to discovery – to more or less anything they wish to market or monetize, including whole books (both print and digital), chapters and other fragments. The advantages of including book content in existing DOI-dependent discovery systems (such as CrossRef) are already being seen. In many cases, the DOI will work alongside and in combination with an ISBN, and many publishers may wish to include an element of intelligence in the otherwise free-style DOI syntax by incorporating the ISBN. One option already being tested in Germany is to create an “actionable ISBN” using a Bookland DOI syntax within the DOI standard, which would allow consistent and unique identification of an ISBN embedded within a DOI, enabling software to infer the required DOI name from an ISBN and vice versa.

In short, the benefits of DOI to book publishers are considerable (as the experience of journal publishers indicates). However, it would be misleading to imply that in any sense the DOI represents an alternative to the ISBN in the trading of digital content.

The International Standard Text Code (ISTC)

The ISTC is an identifier of “textual works”. It does not identify physical or digital products, such as books and e-books, and it plays no role in the trading of content, whether physical or digital. It is intended to facilitate the exchange of information among authors, agents, publishers, retailers, librarians, rights administrators and other interested parties.

One of the most important benefits of the ISTC is that it meets the need experienced by publishers, booksellers, librarians, bibliographic data suppliers and others to collocate different manifestations of the same title. It is also designed to be flexible and granular. It can be used to identify, for example, *War and Peace* as an original work (providing a collocating mechanism for all printed and digital editions, as well as manifestations in a whole range of other media) and it can also be used by publishers to collocate the various editions of a single title or to collocate all the chapters within a single title. This means that it has considerable potential value for managing contracts and royalties as well as assisting the search processes of online retailers, among many other applications.

The ISTC will be launched in 2008 and we recommend that publishers explore the opportunities presented by ISTC for a range of business uses.

Suggestions for best practice

As the new digital supply chain has evolved and in the absence of any standards, publishers have inevitably adopted varying approaches to identifying and describing e-books (which throughout this paper are defined as including both text e-books and digital audio books but not audio books distributed on physical media such as CDs), digital fragments and chapters, and other digital products. It is hoped that this discussion paper will prompt debate and encourage consensus on a standard approach which can be adopted internationally across the industry.

Most publishers have complied with the ISBN standard, which specifies that “each different format of an electronic publication (e.g., .PDF, .HTML) that is published and made separately available shall be given a separate ISBN”. Some, however, have assigned a single ISBN (using the invented and deprecated term “eISBNs”) for use in various ways as a **single** consolidated

internal and/or business-to-business identifier for **all** their e-book formats. Thus the same identifier is used on the publisher’s own site, on third-party e-bookstore sites like eBooks.com, and the sites of library platforms (such as netLibrary, EBL, ebrary). This practice serves only to differentiate the e-book version from the printed version and assumes that the publisher has either no interest in or no need for granular reporting or analysis of any sales by vendor or by format; or that these will be managed by other mechanisms.

This approach is highly unlikely to meet the needs of a proliferating electronic offering in the future and is not recommended. Publishers are strongly advised against the creation and use of “eISBNs”, and the use of the term “eISBN” itself.

Sales of e-books through intermediaries pose distinctive difficulties when, for example, a publisher delivers a file for a title in a single electronic format to a retailer or a digital wholesaler for reprocessing into a variety of tradable formats. In some cases, the publisher may have no interest in what happens to the product once sold to an intermediary. However, there are likely to be situations where the recording of sales by intermediaries does need to be reported back to the publisher for royalty and other purposes.

This situation has been addressed in various ways. Some library e-book wholesalers have chosen to allocate their own ISBNs (with their own prefixes). In some cases publishers do not assign any identifier, and simply pass base files, or even e-book-final-formatted files, to vendors (and particularly to the library e-book vendors) linked to one of the printed book’s ISBNs. This is because some publishers use the hardback or first-allocated/parent print ISBN to manage other internal aggregated systems: contracts management, production ledger, scheduling, rights and royalties. The vendors then assign their own ISBNs to manage their own internal systems, or to report back without confusion to the publisher.

In such situations, it would be good practice for publishers to assign an ISBN to the original digital file they supply and to maintain this strictly for use in internal systems, as an identifier for files delivered to business partners, and as an identifier for those business partners (such as e-book wholesalers, audible.com and similar) to use when reporting to the publisher. In such cases, there would be no requirement or benefit for the ISBN to be made public within the wider supply chain.

Some vendors, however, have allocated their own invented and “ISBN-like” 13-digit identifiers, or adopted the risky strategy of appropriating supposedly available but unused prefixes and ISBN-ranges from other countries. This practice is not acceptable, and publishers should take all possible steps to discourage vendors from adopting it.

It may be in publishers' interests to assign their own ISBNs to intermediaries for each digital format made available. This would guarantee the proper reporting of sales within a framework acceptable to the publisher.

Publishers are advised to evaluate their own commercial and reporting needs for any sale of digital product and to make appropriate identification decisions. The following examples may help in identifying the issues to be considered:

Consumer downloadable e-book

Each format of a consumer downloadable e-book should be assigned its unique, separate ISBN. This is good practice. It enables wholesalers to redistribute the e-book through a variable number of retail outlets. The e-book text will be identical and the functionality largely similar unless the publisher or retailer has designated differential user rights DRM.

Vendor-specific consumer digital content

In the case of vendor-specific e-books, it is good practice for publishers to assign unique, separate ISBNs if they require bibliographic agency discovery and listing, or if they require unambiguous reporting from the vendor. This is particularly desirable if the vendor has additional business models – rentals and subscriptions and print/e-book bundles - which will certainly need to be separated by publishers at the reporting level, at least for royalty payments.

Library e-books

One of the most important issues in the emerging e-book market within the library sector has been cataloging and discovery – with libraries eager to purchase electronic content but unable to get a comprehensive picture of which titles are available within each wholesale environment. As suggested above, publishers may wish to assign separate ISBNs to each of a principal vendors' own versions and hope that the vendors will recognize the value in having their services advertised by this means in the bibliographic databases.

It could equally well be argued that each of the main services provides a different user experience and functionality (some HTML-based, others PDF-based), and that the differences are sufficient to make these e-book offerings distinct products.

Published fragments: chapters and parts

The ISO Standard explicitly allows for ISBNs to be allocated to chapters and parts of monographic publications that are traded separately. To avoid a proliferation of unused ISBNs in the supply chain, it is good practice for publishers to reach early decisions on what level of fragmentation they wish to apply. It is unlikely that publishers will wish to assign ISBNs to all chapters and all parts of all their titles, especially since doing so would place a huge and probably counter-productive burden on the bibliographic agencies' resources and products. Fragments (chapters and parts) should be assigned ISBNs in those cases where any of the three requirements listed previously – trade/discover/report — are important.

Publishers may wish to assign DOIs to chapters and other fragments to assist in discovery. Since one of the principal methods of discovery will be the whole-book text itself, publishers may wish to start including chapter or fragment identifiers alongside the objects themselves in print, online or both versions. This might be analogous to the practice of incorporating a discoverable reference in the running footer of a scholarly journal article. A shorter, human-accessible identifier (ISBN) would be more appropriate in a print book; while DOIs would be more appropriate within digital versions (online or hyperlinked PDF) where there would be no need for transcription. The actionable ISBN concept discussed earlier may be of value here.

Fragments: slice and dice

Digital content (online or downloadable) is readily susceptible to almost infinite fragmentation (potentially down to a paragraph, audio catchphrase, dictionary definition, map, mathematical formula, image or below). While it would be allowed within the ISBN Standard, it would be difficult to see what benefit there would be in allocating an ISBN (or for that matter a DOI) to deeply granulated fragments or to an arbitrary aggregation of fragments (for example, pay-per-view or downloadable access to pages 4, 19 and 114 of a computer manual).

However, publishers may – depending on situations – need to have such sales reported to them, both to satisfy royalty requirements and to provide management information about

users' access to specific items of content. Further work will be needed to evaluate and fulfill this requirement.

Custom publishing

The ISO ISBN Standard states that “a print on demand version of a publication where the content has been specifically customized or personalized for a user and which has limited availability shall not receive an ISBN”. This would include student course-packs – the aggregation of arbitrary fragments which may even cut across different publishers – and it would include digital customization.

Since the trading relationship will generally be confined to one publisher and one consumer, or to one single (usually college) bookstore, identification should be unnecessary. It is therefore good practice not to allocate identifiers since none of the three requirements (trade, discover, report) needs to be met. However, there may be some cases where the physical supply chain, warehousing, shipping, trading, reporting or integrating with the publisher's back-office logistics and financial systems will require an internal ISBN (which should still be a globally unique number although not registered with bibliographic agencies) to be assigned as a physical product identifier.

Subscription-based, licensable and rentable products

Publishers are currently assigning ISBNs and/or DOIs to a wide range of products and services, including subscription- and license-based products, websites and rentals. Since DOIs can equally point to a digital reference (usually a URL displaying a metadata record) to a *physical* product or *digital* resource, publishers may (and probably should) assign DOIs to each and every one of these, resolving either to individual catalog product, information, marketing and ecommerce pages.

Appropriate ISBN allocation is more circumscribed. The ISO ISBN Standard reads: “This International Standard is applicable to monographic publications (or their individual sections or chapters where these are made separately available) and certain types of related products that are available to the public,”, defining publications as being “conceived as a whole in one part or within a finite number of parts ... and made available to the public in any product form”. In this way, the Standard is differentiating finite publications from what it elsewhere calls continuing resources – explicitly excluding publications “issued over time with no predetermined conclusion”, that is, serial publications, journals, open-ended loose-leaf series, and the like, which are usually issued successively with numerical designations, and usually purchased by

subscription rather than one-off. The ISBN Users Manual specifically defines anything subject to continuing update – including online databases and websites – as continuing resources.

It is, however, difficult to be precise with this definition. For example, if a student buys online access to a dictionary that is traded under an ISBN, with or without any right to download, how different is this from a subscription or rental? It might be argued that the differentiators are time (=expiry) and the nature of the implied rights transferred. An ISBN should be used if the implication is that the rights to the item (however much it may dynamically change over time) have been purchased one time only and for perpetuity (or as close to perpetuity as is practicable).

In most cases, publishers' contracts will treat subscription or license deals as secondary rights, paying substantially higher revenue shares to copyright holders. Publishers generally interpret consumer and institutional perpetual e-book sales as direct single-copy purchases, analogous to print sales, though royalty percentages will vary. This makes for very valid reasons for separating these two revenue strands at the reporting level.

Dynamic content

Digital technologies allow for content – whether delivered online, downloaded or printed using POD – to be in a state of continuous revision. This has implications for identification and version control. While the ISBN standard has stipulated that “a separate ISBN shall be assigned if there have been significant changes to any part or parts of a publication”, publishers should follow the basic principle that identifiers will be assigned only where there is a clear and compelling need to do so. In most cases, dynamic changes in digital content will not meet this criterion.

The same principle applies to the updating of files within a POD context. For most changes it will not be necessary or desirable to assign a new ISBN.

Identifiers, ISBN agencies and data aggregators

The supply chain for digital products is still evolving, but one thing is clear: any proliferation of products available digitally, whether complete books in multiple formats or chapters or smaller fragments, poses enormous challenges to existing processes for both publishers and data aggregators. This is partly a matter of volume, but proliferation also challenges existing systems and procedures both in terms of the way product information is delivered and in terms of the way it is disseminated through outbound products.

In the US and UK, the ISBN agencies are part of organizations responsible for collecting product data, so the identification of e-products is also of wider relevance. Since, in most cases, major publishers have large supplies of unused ISBNs, pressure on the agencies – though bound to increase – may be expected to be a gradual process. For data aggregators, however, a decision by a publisher to assign ISBNs to all chapters of all titles, for example, would raise immediate issues and challenges.

This is a matter primarily for the agencies themselves to resolve. However, if proliferation of e-products does materialize, decisions will need to be made in consultation with customers and data providers about a range of issues.

ONIX is reasonably well placed to accommodate digital product in existing data flows, and one of the chief purposes of the planned ONIX version 3.0 is to respond to the challenges of a new supply chain. As publishers depend increasingly on ONIX as a medium for delivering product information, fears about the volumes of data being transmitted are likely to arise again; as will the issue of smaller publishers who do not supply data in electronic form.

Using identifiers for internal purposes

It is inevitable that, as the supply chain for digital products develops, new models will emerge which challenge the recommendations put forward in this paper. There will no doubt be a need to be pragmatic in certain circumstances to resolve specific issues raised by particular trading situations.

In the world of the traditional physical supply chain, publishers and others have adopted identification practices which have not been sanctioned by the ISBN standard or assigned ISBNs to facilitate the operation of various types of internal systems. Inevitably similar needs will arise in the digital world. It is important to be clear that it is generally preferable for these needs to be satisfied by the misapplication of properly assigned and valid ISBNs than by the creation of new identifiers or by the invention of ISBN-like numbers, which could potentially give rise to duplication and confusion in the supply chain.

Publishers should consult their local ISBN agency if they need advice on this issue.

Recommendations

Publishers should use ISBNs to identify all digital products which need to be traded separately.

ISBN assignment should be made on the practical basis that separately tradable items which need to be sold, discovered or reported in the supply chain should be individually and uniquely numbered.

Publishers and others should not use the term “eISBN.” It is ambiguous and is not sanctioned by any existing standard.

Publishers should develop practices and processes for the assignment of ISBNs to digital product that are broadly identical to those applied in the past to printed product and are consistent with the published ISBN standard (ISO 2108).

Publishers should discuss issues of identification and agree on policies with their trading partners where a product is being resold in a different manifestation from that in which it is supplied.

Publishers and others in the supply chain should consider the value of implementing ISTC as a collocating identifier within their operations.

Publishers should consider the value of assigning DOIs as well as ISBNs to digital products they make available, as a means of aiding wider discovery.

Publishers are advised to consult widely through bodies such as the Book Industry Study Group, Book Industry Communication, the Association of American Publishers, the Publishers Association and the International ISBN Agency on practices which they wish to adopt in identifying digital products.

It is essential as the supply chain evolves that standard usages and systems be maintained throughout the industry.

Conclusions and next steps

This document was commissioned to stimulate debate by recommending some principles and suggesting good practice for identifying digital publications in the supply chain. It is not supposed to be prescriptive or the final word on the topic. The Book Industry Study Group and Book Industry Communication are committed to working toward the delivery of clear policies in this area based on the widest participation and consensus. They welcome responses to this

document and to a continuing debate that leads ultimately to policies with which the entire industry can comply.

The traditional supply chain for printed books has benefited immeasurably over a long time from the international acceptance and implementation of clear product identification policies. The emergence and development of an efficient supply chain for digital products demands policies of comparable quality. This document is a first step towards such policies.

About the author

Until his retirement, Michael Holdsworth was Managing Director (UK, Europe, Middle East, and Africa) of Cambridge University Press (2004-2006). As a consultant, Michael now provides strategic advice to UK and US publishing, bookselling, and technology businesses on those digital aspects of the industry which most interest him. He is Chairman of Book Industry Communication (BIC), the UK's ecommerce and standards body for book publishing. He is also Chair of EDItEUR's ONIX International Steering Committee on product information.

Further information

The following web sites contain further information relevant to the issues discussed in this paper:

<http://www.isbn-international.org> (for information on how ISBNs may be correctly applied)

<http://www.collectionscanada.ca/iso/tc46sc9/istc.htm> (for further information on the International Standard Text Code)

<http://www.doi.org/> (for further information on the Digital Object Identifier).

For more information on the identification of digital content and to participate in the debate and the processes that will lead to further work in this area, please contact:

In the United States:

Michael Healy
Book Industry Study Group
Email: michael@bisg.org

In the United Kingdom:

Peter Kilborn
Book Industry Communication
E mail: peter@bic.org.uk

Acknowledgements

The Book Industry Study Group and Book Industry Communication are grateful for the generous support of R.R. Bowker and Nielsen Book in the preparation of this paper.
