

TRANSCRIPT

BISG ANNUAL MEETING OF MEMBERS 2007

November 6, 2007 | 9:00 a.m. to 1:00 p.m.
Yale Club of New York | New York City

.....

Presentation TITLE:

Identifying, Discovering and Trading Digital Publications:
Challenges for the Book Industry

Presentation SPEAKER:

Michael Holdsworth

For most publishers, the impact of the web has been confined to Amazon, online retail, and to online and email marketing – mainly for print books. Up to now, only in the academic and professional sectors has the sale of e-books and digital content achieved any degree of scale. One part of that sector – scholarly journals – has moved over 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.

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, downloaded to a PC or 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 offline consumption.

Publishers have approached this new landscape in different ways. There has been much debate and some genuine differences of opinion among all the interested parties as to what might be best practice, particularly about describing and identifying eBook formats, and about managing chapter and other fragment sales.

There are probably no easy answers. It is in this context that BIC in the UK, and BISG in the USA, working with their international partners and with EDItEUR, the International ISBN agency, and bibliographical services providers, are seeking – as would be expected of them – to provide leadership and direction, and to foster consensus within our industry.

Many new models have been tried and tested within academic and professional. Such models range from the outright perpetual purchase of a digital 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. DRM might constrain or allow the right to move the content between devices (from a PC to a handheld; or to have access on 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 a 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 at one time, or a maximum number within a week or a month or ever. The combinations are almost limitless.

Furthermore, different e-versions of the same base eBook version – in terms of functionality and reader experience – are not the same, any more than, self-evidently, is the digital audiobook version. This is in marked contrast to the world of the eBook. The physical copy of *Harry Potter7*, and the user experience of reading it, is the same whether that copy comes from Amazon or Walmart, Waterstones or Fnac.

eBooks at present broadly divide between paged PDF facsimile displays and reflowable displays specially re-formatted for PCs and handheld devices (such as Amazon's MobiPocket on a BlackBerry, or the Sony e-Reader). Users will need the right version of the eBook for their needs and for their chosen reader – online, or on their PC- or device-based. These different eBook versions do not look or feel the same; some may support additional functions (additional colour, or playing embedded audio, or linking out to the web, or internal hyperlinking) which others do not.

This increase in complexity does not stop there. Digital content is not constrained by the economics of physical eBook ink-on-paper production. Text can be fragmented, combined, re-combined, sliced-and-diced, mashed-up in all manner of new bricolage processes. Content may simply be made accessible online, or made available in custom print form through the new digital POD technologies. Some of these new forms of publishing are, again, familiar from the academic and college sectors – custom-built student course-packs, for example, containing chapters and sections from multiple sources, across multiple publishers.

In fact, viable business opportunities are already being exploited through the sale of more granular fragments – for example, maps, sections of travel guides, and other bespoke subsections – paragraphs or pages.

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 there are available and which they are promoting, ordering, trading and buying. Consumers need to know which versions of digital books are available, and whether these are compatible with their needs or with their software or hardware devices. 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 the market evolves.

There are two established identifiers within the industry – the **International Standard Book Number** ISBN and the **Digital Object Identifier** DOI. They are self-evidently the place to start. And it would clearly

be desirable if, either separately or in combination, these two identifiers could between them be configured or reconfigured to meet at least some of the new requirements.

ISBN and DOI are, of course, complementary, not interchangeable identifiers.

The international book industry has, since it embraced machine communication in the 1970s, managed almost all its identification needs around the *International Standard Book Number* (ISBN). It has been a remarkable success story, as is clear from the fact that 166 countries are now members of the ISBN system.

The principal purposes of the scheme, as defined within the Standard are: to facilitate the trading – ordering and distribution – of books; to allow the compilation and updating of book-trade directories and bibliographic databases; to facilitate the running of retail point-of-sale systems; to facilitate rights management; and to facilitate the accumulation of cross-industry data.

Within most publishers' and booksellers' internal systems, ISBN is also the pervasive de facto key database attribute, tying together sales, distribution, physical logistics, production workflow management, cost allocation, contracts management, rights sales, royalty payment and most elements of financial analysis.

The decision in 2005 to replace its ten-digit form with the ISBN-13 arose from concerns about capacity in the light of the success of the scheme, and – specifically – of the proliferation of new e-publications requiring identification.

While some anxieties remain about numerical bloat within publisher, distributor and bookseller systems, the new prefix has removed the fundamental concern about simply running out of numbers.

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. That challenge needs to be met by metadata models. The choice of identifier is not part of that challenge.

Equally, concerns about the potentially spiralling costs of digital identifiers must be met by the appropriate issuing agencies with imaginative new, and possibly multi-tiered, pricing models.

During the 2005 revision process the opportunity was taken to examine and redefine the scope of material applicable for identification by ISBN. The scope of the new standard for the first time explicitly incorporated electronic publications, chapters of publications, and journal articles.

As identifiers ISBN and DOI have different if complementary functions.

The ISBN is thus unequivocally a supply chain identifier (albeit one conceived in a pre-Internet world), to facilitate discovery, trading and reporting by uniquely *labelling* products – for trading commercially - or for public discovery within a common bibliographical catalogue environment.

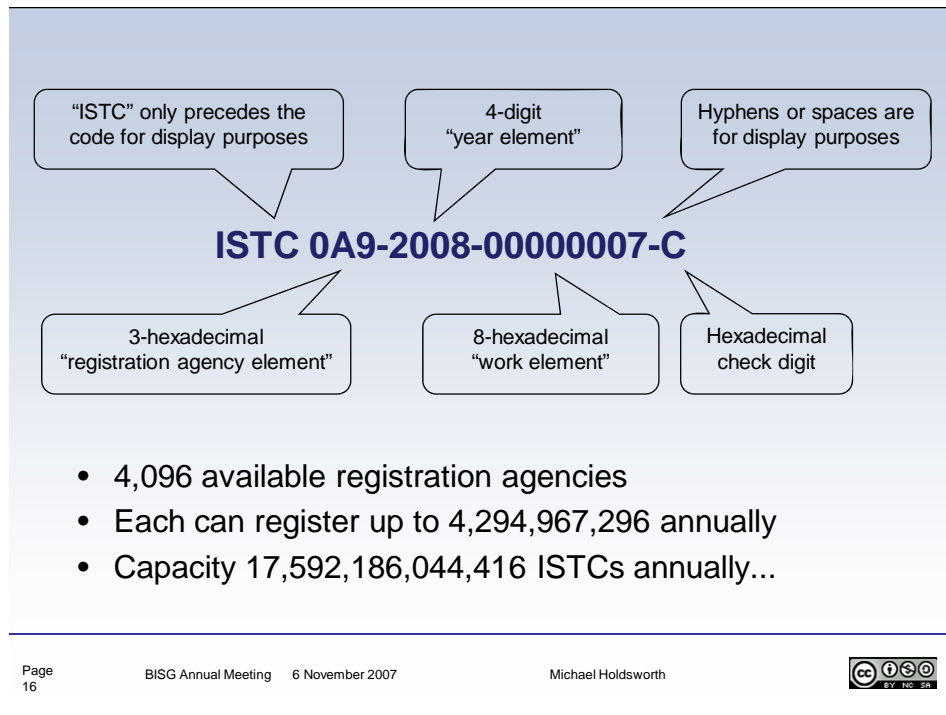
The DOI is a *persistent actionable* identifier which navigates the user to an online web page or pages (with multiple resolution) which can present almost anything – a simple descriptive landing page, or an e-commerce offer, or the resource itself. Information about a digital object may change over time, including its traditional URL or where to find it, but the DOI Name will not change. It is not a

requirement of the DOI that it directs users to a digital product or resource. DOIs can readily resolve to a web catalogue describing a print book. Nearly 30 million DOIs have been issued, principally in support of academic journals.

One of the shortcomings of the ISBN system has always been the problem of consolidation, collocation and archiving – bringing together all versions of a single intellectual textual work under a single identifier. A third, new, identifier - the **International Standard Text Code (ISTC)** which will be launched next year 2008 as the international ISO-approved standard for identification at the ‘work’ level.

Publishers, booksellers, librarians and bibliographical agencies have had their own makeshift solutions to this issue – 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 audiobook etc. (and possibly across translated versions in different languages).

Here’s an ITC – Looks a bit like a DOI, 17 trillion new ones available annually, so no worries about capacity.



Here’s the problem, with everyone’s favourite unread classic Moby-Dick (with its picky-beak hyphen). How to bring all these ‘manifestations’ of the same basic ‘creative work together.

Moby-Dick (or The Whale)

978-0-1234567- 7-X	Penguin Popular Classics
978-0-1234567- 8-X	Wordsworth Classics
978-0-1234567- 9-X	Norton Critical Editions
978-0-1234567- 0-X	Oxford World's Classics
978-0-1234568- 1-X	Barnes & Noble Classics
978-0-1234568- 2-X	Dover Giant Thrift Edition
978-0-1234568- 3-X	Sterling Pop-Up Edition
978-0-1234568- 4-X	Penguin Classics Deluxe
978-0-1234568- 5-X	Longman Critical Editions
978-0-1234568- 6-X	Wiley CliffsNotes
978-0-1234568- 7-X	In Half the Time Compact Editions
978-0-1234568- 8-X	Modern Library Classics

BISG Annual Meeting 6 November 2007

Michael Holdsworth



Of course these are not all identical texts, and ISTC will need to manage that.

Moby-Dick (or The Whale)

ISTC # 1	Penguin Popular Classics
ISTC # 2	Wordsworth Classics
ISTC # 3	Norton Critical Editions
ISTC # 1	Oxford World's Classics
ISTC # 2	Barnes & Noble Classics
ISTC # 2	Dover Giant Thrift Edition
ISTC # 2	Sterling Pop-Up Edition
ISTC # 1	Penguin Classics Deluxe
ISTC # 1	Longman Critical Editions
ISTC # 4	Cliffs Notes
ISTC # 5	In Half the Time Compact Editions
ISTC # 2	Modern Library Classics

BISG Annual Meeting 6 November 2007

Michael Holdsworth



I'm guessing that Penguin and Longman are using the red one – the Oxford text; and that a few of the others – the blue ones, are Gutenberg public-domain versions which won't actually be exactly the same words. And CliffsNotes isn't actually the text at all, so it has its own ISTC.

It's important to note that ISTC won't be a headline identifier of itself, but an attribute of the product ISBN – if you like a field in the metadata or the ONIX.

Moby-Dick (or The Whale)

978-0-1234567-7-X	Penguin Popular Classics
ISTC # 1	Component: OUP text
ISTC # 6	Component: Introduction by A Scribbler
ISTC # 7	Component: Notes by B F Rubbishwriter
ISTC # 11	ISTC for whole Compilation 978-0-1234567- 7-X
978-0-1234567-0-X	Oxford World's Classics
ISTC # 1	Component: OUP text
ISTC # 8	Component: Introduction by B Spears
ISTC # 9	Component: Biographical Note by H M Biographer
ISTC # 10	Component: Essay – “Whales and whaling” by J Starbuck
ISTC # 12	ISTC for whole Compilation 978-0-1234567-0-X



So that one ISBN can obviously contain more than one ISTC, and have its own ISTC for the whole aggregation of contained works, (and which will most of the time be just one work) which will be really helpful for things like royalties and for viewing the stuff from the other direction as here – the basis of an automated work-based catalogue entry.

the work: Oxford World Classics Moby-Dick ISTC 0A1-2007-12B4A105-7 (the compilation) has these identical text ‘manifestations’

print

978-0-1234567- 1-X	hardback
978-0-1234567- 2-X	paperback
978-0-1234567- 3-X	large-print

downloadable eBook *available from multiple retailers*

978-0-1234567- 4-X	Adobe eBook Reader
978-0-1234567- 5-X	Mobipocket

downloadable audiobook *available from multiple retailers*

978-0-1234567- 6-X	Windows Media
--------------------	---------------



As for ISBN, Publishers – as the upstream product originators/owners within the supply chain – determine which of their products will be assigned ISBNs, and they will continue to do so for digital products. I believe that they will do so when in their opinion one or more of the following requirements apply: the requirement for the product to be directly traded; the requirement for the product to be ‘discovered’; and the requirement for transactions to be reported to them

THIS IS THE OLD ISSUE OF EXPEDIENT FUNCTIONAL GRANULARITY – PRODUCTS NEED TO BE SEPARATELY IDENTIFIED TO THE EXTENT THAT THEY NEED TO BE SEPARATELY IDENTIFIED.

DOIs are different – as an entirely optional pointer, they can direct users to any resource, physical or digital, by providing a publisher-controlled and moderated, fully dynamic, and updatable Internet destination landing page from which to provide fuller descriptive information, additional content or ecommerce options.

The DOI has often been described as the ISBN for digital content. This is misleading and unhelpful. It was never intended that DOIs should be used as tradable product identifiers – for example, on orders or invoices or in customer service contexts.

Publishers will increasingly recognise the benefits of assigning DOIs – as an aid to discovery – more or less to everything they wish to market or monetise – to whole books (both print and eBooks) and to fragments. In many cases, the DOI will work alongside and in combination with an ISBN.

The final section of this presentation will look in a bit more detail about the role of ISBN and DOI, will introduce some pointers towards best practice, and will suggest some areas for action. Many issues are genuinely unresolved, and would benefit from further discussion.

First ISBN: The Standard enshrines the principle that individually discoverable, deliverable and tradable product forms require separate ISBNs. It says so pretty explicitly.

ISBN – the Standard

allocation

- *Different product forms (e.g. hardcover, paperback, Braille, audio-book, video, online electronic publication) shall be assigned separate ISBNs.*
- *Each different format of an electronic publication (e.g. '.lit', '.pdf', '.html', '.pdb') that is published and made separately available shall be given a separate ISBN.*

22

BISG Annual Meeting 6 November 2007

Michael Holdsworth



Most publishers have taken this approach with eBooks, allocating separate ISBNs to different formats (for example, Adobe eBook Reader, MobiPocket). Some have gone further, allocating separate ISBNs within the same format to different vendors or platforms (Adobe eBook Reader at netLibrary, or at eBooks.com, or at Ingram MyiLibrary).

I believe that this is likely to prove the only way forward that the supply chain, with its total commercial reliance on a 13 character ISBN will be able to cope with.

Some publishers, 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, and the Sony Connect site, and the eBooks.com site, and the sites of any library platforms (netLibrary, EBL, ebrary).

This practice serves to differentiate the e-Book version from the printed version and it can only be inferred that the publisher either has no interest in or 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. The creation and use of 'eISBN's is deprecated, as is any use of the term 'eISBN' itself.

The same principle of expedient functional granularity will suggest that an ISBN will be required in any circumstances in which the publisher determines that the interests of the product or the supply chain for that product are best served by the assignment of the standard trade identifier. And that it won't, when it isn't sufficiently so served.

For example, if an eBook – either downloadable or serverside – is only available from one site – perhaps the publisher, or netLibrary/MyiLibrary, or Amazon, or Sony or Google – and no-one else can sell that eBook, the publisher may decide not to bother, managing systems and operations with an internal SKU.

Only if the publisher wants the eBook listed in biblio databases would an ISBN provide value. For most other marketing instances, a DOI might do the trick.

Equally, it seems to me that the infinite variations of DRM permission are best offered and managed at the point of delivery, without baking all possible combinations into identifiers – some of which will certainly never be required.

Next: chunking and fragmentation.

The Standard explicitly allows for these: ISBN is applicable to monographic publications (or their individual sections or chapters where these are made separately available)

In many cases, where chapters are only available online (as access or downloads) from just one location, or at least a defined number of locations, DOI may well be sufficient – at least if it is only for discovery. A DOI might resolve simply to the unique selling location; or might resolve back to the publisher. Once in the ecommerce location, transactions can be enabled by clicking on lines or icons which do not carry any labels themselves, but which can locate the appropriate content by internal SKU.

The same principle of expedient functional granularity applies equally here: an ISBN will be required in any circumstances in which the publisher determines that the interests of the product are best served by the assignment of the standard trade identifier.

In short, fragments (chapters and parts) should be assigned ISBNs in those cases where any of the three requirements– trade/discover/report –are important. Initially publishers may only wish selectively to apply ISBNs – perhaps mainly for chapters which appear free-standing, or which are components of multi-author edited works, or conference proceedings.

A similarly flexible approach should probably be taken for custom slice-and-dice content customised on the fly. Digital content (online or downloadable) is readily susceptible to almost infinite fragmentation (potentially down to a paragraph, audio catchphrase, 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 ISBN (or for that matter DOI) allocation to a one-off deeply granular fragment; or to an arbitrary aggregation of fragments (for example, pay-per-view or downloadable access to three pages - say 4, 19 and 114 - of a computer manual). The same would be true for the custom publishing of course-packs – for identification either at the pack or component level.

Finally a word next on digital products which aren't traditional 'books'. Publishers are currently assigning ISBNs and/or DOIs to a wide range of products and services, including licences, subscriptions, 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 catalogue product, information, marketing and ecommerce pages.

Appropriate ISBN allocation is more circumscribed. The ISO Standard differentiates 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 purchased on subscription rather than by one-off purchase.

It is likely that best practice will determine that any arrangements which manage licensing deals, or require regular payment or renewal should be considered ineligible for ISBNs. This includes arrangements (perhaps for bundled access to content on a college textbook website; or Google rentals) which will carry a fixed expiry date.

So where does all this leave us?

The track is now pretty much laid for the standards for eBook digital content itself. Amazon, Google may stay with some proprietary formats for perceived competitive advantage, and Sony has compromised, but for the rest, hard graft and cooperation choreographed by the IDPF and Adobe for .epub and PDF standards has finally paid off.

The identifiers agendas?

First – for the industry bodies, working with the bibliographical services providers and their international partners, to attempt a codification of use-cases and best practice for the use of ISBNs and DOIs within the new digital supply chain. This work is underway. We require proselytisation and evangelism for wider usage and for the correct wider usage of the available standards.

use-cases

downloadable eBooks
audiobooks
online access
print/eBook bundles
Google and Amazon
POD and 'live books'
custom publishing
licences, subs, websites and rentals,
chunks, fragments and articles

Second – for some real and practical progress to be made by national standards groups everywhere in the roll-out of ISTC. So that we can have web catalogue pages structured like this.

**the work: Oxford World Classics Moby-Dick
ISTC 0A1-2007-12B4A105-7 (the compilation)
has these identical text ‘manifestations’**

print

978-0-1234567- 1-X	hardback
978-0-1234567- 2-X	paperback
978-0-1234567- 3-X	large-print

downloadable eBook *available from multiple retailers*

978-0-1234567- 4-X	Adobe eBook Reader
978-0-1234567- 5-X	Mobipocket

downloadable audiobook *available from multiple retailers*

978-0-1234567- 6-X	Windows Media
--------------------	---------------

BISG Annual Meeting 6 November 2007

Michael Holdsworth



**Oxford World Classics Moby-Dick
ISTC 0A1-2007-12B4A105-7 (the compilation)
has these identical text ‘manifestations’**

eBook/audiobook *vendor-specific*

978-0-1234567- 7-X	downloadable Sony
978-0-1234567- 8-X	perpetual online Google
978-0-1234567- 9-X	perpetual online Amazon
978-0-1234567- 0-X	perpetual online Publisher’s own site
978-0-1234568- 1-X	print/online Amazon Upgrade
978-0-1234568- 2-X	downloadable audible.com
978-0-1234568- 3-X	downloadable iTunes

BISG Annual Meeting 6 November 2007

Michael Holdsworth



**Oxford World Classics Moby-Dick
ISTC 0A1-2007-12B4A105-7 (the compilation)
has these identical text ‘manifestations’**

library eBook *vendor-specific; ‘tethered’ access/downloadable*

978-0-1234568- 4-X	Overdrive DL Reserve
978-0-1234568- 5-X	netLibrary
978-0-1234568- 6-X	Ingram MyiLibrary
978-0-1234568- 7-X	ebrary
978-0-1234568- 8-X	Dawson
978-0-1234568- 9-X	Blackwell

BISG Annual Meeting 6 November 2007

Michael Holdsworth



Third – for the bibliographical services companies and issuing agencies to get to grips with the inclusive delivery of information feeds about digital content to their bookseller and library customers who have for too long been denied this; to build out new services based on DOI and ISTC; and to explore new flexible models for identifier pricing.

Fourth – for EDItEUR’s ONIX groups to resolve outstanding issues about eBook and digital content product description. Here are some pointers to the new fields we might need.

ONIX – eBook add-ins

product form (Mobipocket, 'VitalBook, AER)
images present or not
reflowable or not
e-pub package
encoding format version
display height/display width
image resolution
file size
video or audio bit-rate
DRM values

34

BISG Annual Meeting 6 November 2007

Michael Holdsworth



Fifth – for the ONIX groups to get guidance out there on the way the standard will work within increasingly granular environments – from parts, to chapters to paragraphs to images, maps, tables. Stuff like this.

ONIX – fragment add-ins

component descriptive blurb
component level TOC
component author
place in serial series (as in Chapter **one** of eight)
number in total series (as in Chapter one of **eight**)
component title, subtitle etc.
print start and end page numbers
essential non-revenue apparatus
– identifiers for other components which would always ship with this fragment - notes or bibliography section, maps section, index, front-matter etc.

36

BISG Annual Meeting 6 November 2007

Michael Holdsworth



Sixth – that new work be started on consistent naming of marketing files – perhaps following the lead provided by Amazon submission and the BATCHConnect initiative in the UK. Although I haven't shown it

like this here, there would be considerable value in collocating these under an ISTC umbrella. It's not just file-naming; it's also identifying; and designing a metadata dictionary.

Finally, seventh – to progress valuable work (piloted by MEDRA) on the extent to which some ISBN intelligence might be appropriately built into a new hybrid ISBN-DOI-co-standard – the so-called Actionable ISBN or 'Bookland' DOI

