



Apex CoVantage, LLC



Using DOIs for Books in the Mobile Environment

A Report Prepared for CrossRef
by Bill Kasdorf, VP—Apex Content Solutions

June 30, 2012

Revision History

Rev Num	Release Date	Changes	Author
1.0.b1	February 11, 2012	Initial Draft	Bill Kasdorf
1.0.b2	March 19, 2012	Revised based on initial feedback, esp. re DOI display guidelines, and drafting Resources section	Bill Kasdorf
1.0.b2	June 30, 2012	Revised to eliminate recommendations regarding work identifiers	Bill Kasdorf



Contents

1.0	Different Identifiers for Different Purposes.....	1-1
2.0	The Digital Object Identifier.....	2-3
3.0	Identifiers and eBooks	3-7
4.0	Recommendations for the Use of DOIs for Scholarly Books in the Mobile Environment.....	4-14
5.0	Resources	5-18

1.0 Different Identifiers for Different Purposes

There is a great deal of confusion today regarding the use of identifiers for eBooks. The most widely used identifier, the ISBN, is often used incorrectly; the ISTC, an identifier designed to address the need for a “Work” identifier, is seldom used; and other identifiers are used for various types of books, components of books, or for various purposes in editorial, production, manufacturing, and distribution workflows.¹

As eBook platforms have proliferated and as the demand to access eBooks on mobile devices has become commonplace, the need for practical, workable, unambiguous identification of eBooks has become urgent. Publishers need to keep track of the various distinct eBook versions they offer while not losing the ability to associate them all with a single title, to which they need to attract customers and users; users need to find eBooks and obtain the versions or formats—or even the portions of those books—that they seek; retailers, distributors, aggregators, and librarians need to both distinguish and associate all of these possible variations; and students and scholars need to be able to cite and link to eBooks, ideally at a very specific, granular level. This is what identifiers are for.

One of the ironies of book publishing today is that the identifier that is arguably the most useful of all in addressing these issues—the DOI or Digital Object Identifier—is currently seldom used by book publishers, despite the fact that its development was initiated in 1996 by the Association of American Publishers (AAP), focused largely on books.² Now an international standard,³ the DOI is most widely used to identify scholarly journal articles; in book publishing, it is used mainly by the publishers of scientific, technical, and medical (STM) books who also use the DOI for journal articles and make book content available online. This report is being published by CrossRef specifically to guide the use of DOIs for scholarly books in the mobile environment, but it is hoped that it will help encourage the use of DOIs more widely in book publishing in general.

It’s important to realize that the DOI should be used *in addition to*, not instead of, other identifiers. The two most relevant to this discussion are the ISBN and the ISTC.

The ISBN: For Products

The International Standard Book Number (ISBN) is fundamentally a *product* identifier for the supply chain. It is used to identify each distinctly different version or format of a book. While publishers have almost always used it properly in the context of print and audio books, issuing separate ISBNs for the hardcover, softcover, and audiobook, they often incorrectly assign a single ISBN to all eBook versions, or sometimes one to the PDF and one to all reflowable eBooks. This undermines the purpose of the ISBN: the eBooks version won’t work on the Kindle, and reporting sales information based on a single ISBN means

¹ The most comprehensive overview is the “Roadmap of Identifiers” published and periodically updated by the Book Industry Study Group (BISG), available at http://www.bisg.org/docs/Roadmap_of_Identifiers.pdf.

² The DOI was ultimately developed and is governed by the International DOI Foundation (IDF), <http://www.doi.org>, which was formed in 1998 by the Association of American Publishers, the International Publishers Association, and the International Association of Scientific, Technical and Medical Publishers (STM).

³ The DOI was approved as an ISO standard on November 10, 2010 and is in the process of being published as ISO 26324; it complements ANSI/NISO Z39.84-2005, *Syntax for the Digital Object Identifier*.

that eBooks and Kindle sales can't be distinguished by ISBN. The ISBN is not a "work" identifier, and there is no such thing (officially) as an eISBN.

The ISBN is governed by the International ISBN Agency (www.isbn-international.org). ISBNs are issued by national ISBN agencies such as Bowker in the US (www.isbn.org) and Nielsen in the UK (www.isbn.nielsenbook.co.uk). The ISBN is a 13-digit number that identifies the national agency that issued it, the entity (usually but not always the publisher) that obtained it, a unique string of arbitrary numbers, and a check digit used to ensure validity of the ISBN. A publisher can't "create" or "make up" ISBNs; they can only be issued by the appropriate agency, who provides each client a specific set of ISBNs to assign, one-by-one, to their various products.

For a thorough discussion of proper use of the ISBN, particularly relating to eBooks, see "Best Practices for Identifying Digital Products," published by the Book Industry Study Group (BISG) on December 7, 2011, at http://www.bisg.org/docs/BISG_Policy_1101.pdf.

The ISTC: For Textual Works

The International Standard Text Code (ISTC) is intended to uniquely identify a textual work,⁴ independent of the various manifestations (editions, versions, formats) in which it is published. In fact, its main purpose is to provide a means for associating the various editions, versions, and formats that are "the same content." Because each of those products is assigned a separate ISBN, and because there is no way to infer from one ISBN the ISBNs of the other versions, assigning all versions the same ISTC (in addition to their individual ISBNs) enables them to be associated.

The ISTC is governed by the International ISTC Agency (<http://www.istc-international.org>) and issued by one of a number of Registration Agencies (ISTC RAs). At the time of this writing, there are ten ISTC RAs (including Bowker and Nielsen); in contrast to the ISBN, there is no requirement, geographical or otherwise, for use of a particular ISTC RA.

The ISTC is a hexadecimal (16-digit) number that consists of a registration element, the year the ISTC was registered, the work element, and a check digit. Once assigned, it is intended to permanently identify the work with no implication as to publisher, version, or format. Specific metadata is required to register an ISTC; this metadata serves both to distinguish works that might appear to be the same (e.g., a novel and a play with the same title by the same author) and to prevent different ISTCs from being assigned to the same work (e.g., by two publishers each attempting to register the same out-of-copyright work).

For a thorough discussion of the issue of work identification (and the related issue of component identification), see the June 2011 discussion paper "Requirements for the Identification of 'Components'" by Mark Bide and Graham Bell of EDItEUR available at <http://www.editeur.org/117/Publications/>.

⁴ It should be noted that the precise definition of "work" differs in different contexts. There are two main abstract models in general use. The ISTC and the DOI are based on the <indecs> model, which has a focus on rights and commerce (see http://www.doi.org/topics/indecs/indecs_framework_2000.pdf for a detailed discussion of "The <indecs> Metadata Framework"); the model used most often in the library world for cataloging and related bibliographic applications is FRBR (see <http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records>).

2.0 The Digital Object Identifier

The DOI is a completely different kind of identifier: it's an *actionable* identifier. When a user clicks on it, something happens. And its registrant controls what that "something" is.

Especially in the context of books and mobile delivery, it's important to realize how fundamentally different from the ISBN and ISTC the DOI is. The ISBN and ISTC are *received* identifiers: they can't be created, in any respect, by their owners; you are issued ISBNs and ISTCs, you can't invent them. And their scope is very specifically defined and limited: the ISBN is *only* a product identifier, and the ISTC is *only* a textual work identifier. While metadata must be supplied to characterize what they identify, their use is in a sense quite independent of that metadata: while an ISBN or an ISTC can potentially help lead to information about the product or work they identify, there is no universal system or infrastructure that does that.

The DOI is different in all of these respects.

- Although the DOI, once assigned, is a "dumb number" that is not intended to imply anything about the object it identifies, the registrant of a given DOI has a great deal of control over what makes up that "dumb number" when it is created in the first place. As we'll see below, it is possible—and common—to embed other identifiers in a DOI. Even an ISBN or an ISTC.
- A DOI can be assigned to a book or its content at any level of abstraction: at the work level, at the product level, at the component level, and at virtually any granular level within the content. In fact, DOIs can be registered for any or all of these for a given book.
- Most importantly, the DOI provides resolution services to its owner, establishing a permanent, persistent identity that enables its Registration Agency (RA) to provide flexible, dynamic services such as reference linking or redirection to changing URLs.

Let's take a look at how this works.

The DOI: Actionable and Adaptable

The first misconception about the DOI to clear away is that it is only for digital content. It's a *digital identifier* of objects (physical or digital), not an identifier of digital objects only. It's as relevant to print books as to eBooks.

Its fundamental purpose is to provide a persistent, interoperable, actionable link to whatever its owner wants it to link to. Unlike URLs, DOIs never change: once they are assigned to something, they always and forever identify the same "object."

However—and this is where the most useful and dynamic functionality comes in—not all the metadata associated with a given DOI is permanent. The DOI system is designed to enable the metadata to change. That means that a DOI assigned to a given book is always associated with that book, even if the party who controls the DOI changes (for example, when a book's publisher is acquired by another publisher), in which case the URL to which the DOI resolves is only one of many important aspects of metadata that needs to change—without having to change the DOI and without all the DOIs that are "out there in the wild" ceasing to

work. Even without such a drastic change, a publisher needs to update what the DOI links to from time to time: perhaps initially just to a publisher’s website or online catalog; later to the URL of the page about that book; or perhaps to an intermediate dialog that presents a choice of options: places to buy the book, available versions, information about the author, etc.—whatever the owner of the DOI wants it to do and its Registration Agency enables. And publishers need add metadata over time as well: perhaps additional identifiers like ISNIs for organizations or ORCIDs for authors, or awards and excerpts from reviews that help sell a book.

The relationship with a particular DOI Registration Agency (RA) cannot be overemphasized. The services associated with a DOI are provided by its RA; while it is possible to transfer control of a DOI from one RA to another (because the DOI itself cannot change, once assigned), this rarely happens. Instead, publishers use specific RAs to obtain specific services (which of course the RA can modify over time). As an example familiar to most readers of this report, CrossRef—the RA that at this time has assigned the vast majority of DOIs—uses the DOI primarily to provide reference linking services for scholarly content. DataCite is an RA that uses the DOI to provide links to sets of scientific data. The Publications Office of the European Union (EU), another RA, uses the DOI to provide links to all official EU publications.

Deconstructing the DOI

The DOI itself, like the ISBN and ISTC, is an alphanumeric string, but it differs in two important ways: its length is not prescribed, and it enables the registrant to determine much of its content. It is composed of two parts, the prefix and the suffix, separated by a forward slash (/), e.g.:

10.1234/xyx56789abc
or
10.1234/MyBookChapter5

and would be expressed online (and, ideally, even in print)⁵ as hypertext links like this:

<http://dx.doi.org/10.1234/xyx56789abc>
or
<http://dx.doi.org/10.1234/MyBookChapter5>

where dx.doi.org is the DOI proxy server.

The prefix begins with “10.” followed by digits that are assigned by the RA to the registrant. (Note that an organization can be assigned more than one prefix, which can

⁵ CrossRef strongly recommends always expressing DOIs in this full URL form. (See CrossRef DOI Display Guidelines, http://www.crossref.org/02publishers/doi_display_guidelines.html, for a very thorough set of recommendations in this regard.) Publishers sometimes question why this is appropriate even in print. The two most important reasons are that modern workflows typically repurpose a given set of textual content for delivery in multiple modes; if that content does not have DOIs in this recommended URL form, they will not be live links when that content is published online. Also, publishing the full form enables users to copy and paste them in their most useful form. (It should also be noted that this recommendation to express DOIs as full URLs is a relatively recent change from CrossRef; many style guides still reflect the earlier recommendation simply to use the doi: prefix.)

help administratively when different groups of publications—e.g., imprints, or books vs. journals—need to be distinguished.) The registrant may then provide any alphanumeric string (DOIs are case-*insensitive*) following the slash to identify the object being registered, as long as that string is *unique to the registrant*. (The prefix is what makes it unique across all DOIs, thus eliminating the need for the suffix to be universally unique.)

This is what gives the DOI such flexibility. The suffix can be an arbitrary ID meaningful to the publisher; a simple sequential number; an ISBN (commonly done; but this should only be done when the DOI refers to the *product* designated by that ISBN); an ISTC (again, ideally only when the DOI refers to the textual work, rather than manifestations or products of it); or whatever else the registrant finds appropriate.

Once the DOI is assigned, it permanently identifies the registered object. It becomes a “dumb number”: no ownership or rights information can be inferred from it, and neither can the nature of the object it identifies be inferred as work, product, or component. Specifically, the digits between the period and the slash in the prefix no longer “mean” the publisher or other original registrant. The DOI of a book registered by Hungry Minds using the Hungry Minds prefix did not change when Hungry Minds was acquired by Wiley. And if Hungry Minds used an ISBN as the suffix to refer to “the book,” that suffix doesn’t change when Wiley subsequently publishes the book in eBook formats that have separate ISBNs. Wiley can register additional DOIs, of course; but the original DOI never changes.

DOIs in Action

Why is this such a good thing? *Because the metadata associated with the DOI—including, most importantly, a URL to which it links—can change.* Wiley can submit to the RA new metadata that provides the URL of its own catalog page for that book it acquired from Hungry Minds, replacing the previous Hungry Minds URL (*redirection*), and can even provide metadata about all the new formats in which the book is now available, and how to get them (*multiple resolution*).

This is all managed through a very robust infrastructure called the Handle System. The Handle System is the key to the provision of services associated with DOIs. Although each Registration Agency provides specific services geared to the community it serves, based on specific metadata that it requires of registrants, all DOIs are resolved through the Handle System, which links the DOI to the web location controlled by its registrant and managed by its Registration Agency.

The services provided by CrossRef include both *reference linking* and *multiple resolution*. Thus a book publisher could register:

- one CrossRef DOI for a book at the title level (perhaps by using an ISTC as the suffix, e.g. <http://dx.doi.org/10.1234/ISTC-0A9-2012-12B4A105-7>), which would be the most useful for citation purposes;
- one CrossRef DOI for each chapter of that book (e.g., http://dx.doi.org/10.1234/ISTC-0A9-2012-12B4A105-7_Ch5 or <http://dx.doi.org/10.1234/MyBookChapter5>), which would facilitate the citation or sale of individual chapters;

- one CrossRef DOI for a component of the book (e.g., http://dx.doi.org/10.1234/5MCC_AlzAlg), which is especially valuable if that component, such as a diagram or an image, has value and utility apart from the book;
- one CrossRef DOI for each eBook version (e.g., <http://dx.doi.org/10.1234/ISBN-978-0-123-45678-9>), which would facilitate eCommerce and distribution—including discovery of the *other* versions.

No publisher is likely to issue all of these CrossRef DOIs for all of their books; but these different uses are relevant, in various combinations, for most publishers' books.

The resulting CrossRef DOIs can, and should, be disseminated as widely as possible. The chapter-level CrossRef DOI would be ideal to embed in a citation of that chapter in another book or other publication (e.g., a journal article citing that chapter); the title-level CrossRef DOI would be ideal to embed in a book review or in a library catalog (whether institutional or personal); the CrossRef DOI for an iPad app would be ideal to embed in a review of that app, or even in an e-mail or blog post recommending that app. And the publisher could use any or all of those CrossRef DOIs not only to direct the user to exactly what that CrossRef DOI is registered for (the chapter, information about the book in general, the iPad app) but also to provide, or direct the user to, information about that book or about related books or other resources.

A Note on ISBN-A

It should be noted that a special-purpose type of DOI called ISBN-A, an “actionable ISBN,” has been recently developed.⁶ While that may appear to be a special purpose type of ISBN, it is really a method for embedding an ISBN in a DOI. It does not replace the ISBN; instead, it incorporates a given ISBN's digits in DOI syntax and must be registered as a DOI. This can only be done by a DOI RA that is also an ISBN RA, because the services provided are both DOI- and ISBN-related. In the ISBN-A syntax, the portion of the ISBN designating the registrant becomes part of the DOI prefix, and the portion identifying the title becomes the suffix, e.g. 10.978.12345/99990.

Note that this does not preclude the use of the ISBN in the suffix of a “normal” DOI, e.g. “10.1234/ISBN9781234599990,” which can be registered by any publisher through any DOI RA. But that does not in itself imply or enforce ISBN-related validity; in that case, the publisher just used the ISBN in the suffix of the DOI for its own reasons. The ISBN-A, on the other hand, does formally combine the DOI and the ISBN, which is why the services associated with it need to be controlled by a single RA.

⁶ See “The ISBN System in Relation to the DOI® System,” a DOI System Factsheet available at <http://www.doi.org/factsheets/ISBN-A.html>.

3.0 Identifiers and eBooks

One reason the issue of standardized identifiers for eBooks has been so murky is that until recently there didn't seem, to most publishers, much need for them except in a production workflow and perhaps to some extent in the supply chain—both of which were not only tolerant of, but most often engineered for, proprietary identifiers. The ISBN seemed to be sufficient (and as noted above was often used incorrectly in the context of eBooks). Digital books—except for reference books and some STM titles—were rarely made available for direct access online; instead, they were (and still are) normally delivered as complete, self-contained files (whether as PDFs, EPUBs, apps, or other formats): *products* that are discovered, sold or licensed, and delivered either through commercial channels or by libraries, often for offline reading, or at least not requiring a reader to be online. The ISBN, as a product identifier, was sufficient for eCommerce, and libraries use separate cataloging and identification that does not originate with the publisher. Because books were not typically read online, there was little demand for reference linking; and even when books *were* read online, the likelihood that a cited book would be able to be linked to was small.⁷

With the proliferation of eReaders, smartphones, tablets, and the mobile environment, that has all changed. The lack of standardized eBook identification has become a problem for everybody—publishers, retailers, aggregators, libraries, and readers. Without good identification mechanisms, discovery is impaired, access is limited, and citation is impossible. The DOI, when used well, can address all of these issues. Its permanence, its flexibility (both in regard to its content and in what it identifies), and the services that can be associated with it all distinguish it from other available identifiers. Recommendations for its use in eBooks are given in the section that follows. First, some background on identifiers in eBooks in general.

Identifiers are part of the metadata associated with a book. As such, they can be delivered both independently of the book (as in the metadata feeds publishers commonly provide to retailers and aggregators) and embedded in the book itself.

Identifiers in External Metadata

Metadata is provided to the supply chain by a number of different mechanisms. The most common are spreadsheets and as XML, ideally in the form of ONIX metadata. Many of these metadata feeds convey title information in the commonly used Dublin Core⁸ vocabulary (e.g., “dc:identifier” for an ISBN or DOI), although more are still in proprietary or unspecified syntax.

While it would be desirable for both the publishers and the recipients of external metadata to use well established standards, the opposite is currently the case. Although good

⁷ The exception, of course, is CrossRef. As of this writing, the CrossRef database has 260,823 books and 4,427,317 book DOIs (thus almost 17 DOIs per book, indicating a large number of chapter and component DOIs). But this is still a very small portion of the books—even in the scholarly realm, to say nothing of textbooks, trade books, professional and reference books, etc.—that could and should be citable and linkable.

⁸ The Dublin Core Metadata Element Set (DCMES), which is extremely flexible and tolerant, is most commonly used; another form of Dublin Core metadata designed for more explicitness needed for “linked data” is dc:terms. See <http://dublincore.org>.

standards exist—MARC records in the library world and ONIX in the commercial supply chain being the two most prominent—they are currently not used consistently and are often not used at all by publishers. MARC records are typically created by catalogers rather than being generated from the publisher’s metadata. ONIX is an extremely sophisticated and accommodating standard for book supply chain metadata, but its very richness leads to inconsistent use on the part of publishers; and the transition to the current version, ONIX 3.0, which has significantly improved capabilities for eBook metadata, has been slow. All too many publishers still submit their metadata as spreadsheets.

Even more problematic is the fact that retailers, aggregators, and distributors—from the dominant Amazon and Apple to the hundreds of other such players in the supply chain—typically require a specific, proprietary set of metadata for books. Compounding the problem is that the databases most publishers use to manage their metadata were created before the eBook era and are difficult to adapt to today’s needs. Finally, much eBook metadata is currently not even created by the publisher; instead, its creation is commonly done by third parties (such as the vendors creating the eBooks) and is often inconsistent with the metadata for the print versions of the same books.

The result: publishers complain that they have to provide metadata in too many different forms meeting too many different specifications; the recipients of that metadata complain that they can’t conform to a standard because the metadata from publishers is too inconsistent. It is an unfortunate tower of Babel.

DOIs Even Work in Bad Systems

The good news, however, is that virtually all of these metadata schemes and delivery mechanisms can accommodate the DOI. While few of them require it, and none use it in any systematic way, it is possible to provide or incorporate a CrossRef DOI in virtually all of them, whether as the only identifier for a given title, or as one of many. (In all cases, but particularly the latter, it is important to express the CrossRef DOI in its citable form, e.g. <http://dx.doi.org/10.1234/xyx56789abc>, which clearly and unambiguously identifies it as a DOI, because not all systems provide the ability to identify the *scheme* of an identifier.)

In general, that CrossRef DOI should be a title-level DOI rather than being a DOI associated with a specific product or component. The reason is that for external metadata—that is, metadata that exists apart from the book or books it describes—it is most useful to both the publisher and the ultimate end user if that CrossRef DOI presents or links to information about the title in general, rather than specific manifestations of it. (That can of course be done with any DOI; but it is normally a title-level DOI that is used for this purpose.)

Most importantly, this use of the CrossRef DOI in external metadata can serve to liberate the publisher from that tower of Babel. No matter what metadata is in a given retailer’s, aggregator’s, or distributor’s system (controlled by them), the CrossRef DOI provides a way to link to whatever metadata the *publisher* wishes to present, via its RA. It is also transferable to other systems, other publications, even reviews and e-mails—and in all those cases it is reliable, citable, and *actionable*. No other identifier can do that.

Embedding Identifiers in eBooks

Identifiers can be embedded in eBooks for three distinct purposes:

- To identify the eBook itself.
- To identify components of the eBook.
- To identify something cited in the eBook.

Most current eBook formats accommodate the incorporation of identifiers but do little to enable their management and use in systematic ways.

- The PDF format can include “Document Properties” where a URL or a “custom property” could be provided, and it also accommodates XMP⁹ metadata, which has an identifier component; while these are rarely used, it would be possible to embed a DOI in them.
- EPUB 2.0.1, the current most commonly used standard XML-based general-purpose format for reflowable eBooks, contains an <identifier> element at the “package” level that is specified as “A string or number used to uniquely identify the resource.”¹⁰ (The “package document,” with the extension .opf and the root element <package>, is the component of an EPUB that contains its metadata.) EPUB 2.0.1 can accommodate more than one <identifier>, but one of them must be identified by the @unique-identifier attribute associated with the <package> that uniquely identifies that specific EPUB. In addition, an optional @scheme attribute is available for <identifier> which, to quote the EPUB 2.0.1 spec as cited above, “names the system or authority that generated or assigned the text contained within the identifier element, for example ‘ISBN’ or ‘DOI.’ ” While this is rarely done in most of today’s EPUBs, ideally both the ISBN and the DOI should be provided. Because an EPUB can be both a *product* provided to end users and a *source* from which subsequent products are generated, a DOI registered specifically for that EPUB would be the ideal “unique identifier” for it (rather than one of several ISBNs) because it could link to both the general-purpose EPUB and the individual products created from it.
- As mentioned above, eBook aggregators, platforms, and reading systems typically have their own proprietary specifications for metadata. While most of them don’t require the DOI or use it systematically, most can accommodate it within their metadata at the title level. Few enable its use at a chapter or other component level, except the leading STM online hosts, who do commonly use chapter-level CrossRef DOIs.
- Most eBooks provide the ability to link to an external resource on the Web if the user is online. This means that CrossRef DOIs could be embedded in them as links—for example, in citations—in the form <http://dx.doi.org/10.1234/OtherBookChapter5>. The link wouldn’t necessarily have to resolve to the actual cited content; the owner of the cited CrossRef DOI controls that, so if the cited content itself is not available online, the CrossRef DOI could direct the user to a publisher’s catalog, to a particular catalog page, or to purchase or other access options.

⁹ XMP is Adobe’s “Extensible Metadata Platform.” It is commonly used to embed metadata in graphic files like PDF, Photoshop, and Illustrator documents. See <http://www.adobe.com/products/xmp/overview.html>.

¹⁰ See http://idpf.org/epub/20/spec/OPF_2.0.1_draft.htm#Section2.2.10 for details on <identifier> in EPUB 2.0.1.

A Word about the shortDOI

Publishers often find the DOI—especially when properly expressed as a URL—to be cumbersome in some contexts because of its length. This is mostly an issue for DOIs in print, where real estate is at a premium and where long DOIs can create line-break issues. It is especially a concern in reference lists, where there can be a great number of DOIs.

One solution is to obtain a shortDOI,¹¹ a service developed by the IDF to create a short handle in the form 10/abcde for a longer DOI. This shortDOI functions exactly as its longer counterpart does, but takes up much less space. Anybody can obtain a shortDOI by submitting a long DOI, either for their own DOIs or the DOIs of other publishers' publications that they're citing. If a shortDOI has already been issued for a given long DOI, it will be returned for every request relating to that DOI, thus ensuring that the same shortDOI is used everywhere, once issued, for a given long DOI.

Identifiers in EPUB 3.0

The latest version of the EPUB specification, EPUB 3.0, provides dramatically enhanced metadata capabilities. Whereas EPUB 2.0.1 only provides for metadata at the package level, EPUB 3.0 provides for the use of greatly expanded metadata at the package level, and also enables the incorporation of metadata about items at the component level (for example, on individual chapters or on other resources such as images), and even at a very granular level within the HTML5-based content documents. In addition, EPUB 3.0 enables publishers to include metadata resources like an ONIX file or a MARC record within the EPUB.

Although EPUB 3.0 *enables* publishers to include much more metadata, it doesn't *require* much more than EPUB 2.0.1. Thus it accommodates any number of identifiers in the package metadata (in the <dc:identifier> element, using Dublin Core); it enables but does not require the identifier scheme to be specified (using an "identifier-type" property, with an optional @scheme attribute); and it requires one identifier to be designated the "unique identifier" of that EPUB via the @unique-identifier attribute on the top-level <package> element.¹²

But there is one very important difference in how EPUB 3.0 treats these identifiers. Although EPUB 2.0.1 used the term "unique identifier," in practice there was variation in the interpretation of the word "unique." Technologists tend to use that term very literally: if anything at all is changed in an EPUB (even the addition of a comma), the new EPUB should get a new unique identifier because the two EPUBs are not absolutely identical. But publishers tended to take a more practical view, overlooking small changes like fixing typos, viewing the updated versions as basically "the same EPUB" and thus not changing the unique identifier.

¹¹ For more information on the shortDOI, go to <http://shortdoi.org/>.

¹² See <http://idpf.org/epub/30/spec/epub30-publications.html#sec-opf-dcidentifier> for the exact specifications for identifiers in EPUB 3.0, which includes an example of specifying a DOI. This specification has been developed to make the identifier as "machine readable" as possible. While it requires reading systems to infer the identifier type if it is not supplied (which of course is greatly aided by expressing the DOI with the <http://dx.doi.org/> prefix), it also provides the ability not only to explicitly identify the identifier type but also to provide a machine-accessible authority for what that means. The example in the EPUB 3 specification uses ONIX Code List 5 with a value of "06" to unambiguously specify the DOI.

EPUB 3.0 was designed to address both the need for persistence (“it’s the same publication”) and accuracy (“it’s the identical file”) by doing two things:

- It specifies that the “unique-identifier” should be *persistent*: “The Unique Identifier of an EPUB Publication typically should not change with each minor revision to the package or its contents, as Unique Identifiers are intended to have maximal persistence both for referencing and distribution purposes.”
- It creates a new requirement to add a “modified” property (dcterms:modified) in the package metadata that contains a timestamp recording when *that file* was made.

This simple solution addresses both needs. By being persistent, the “unique-identifier” serves as a *publication identifier*, only changing with what is essentially a new publication (or a new edition).¹³ But when combined with the required timestamp, the pair serves as a *package identifier*, addressing a reading system’s need to differentiate versions.

From the point of view of the DOI, it becomes obvious that the CrossRef DOI is ideal as the publication identifier—in this case, specifically the *EPUB publication*. Even if various products are derived from that EPUB (and thus get individual ISBNs), they can all maintain their family identity via the CrossRef DOI. And the CrossRef DOI’s multiple resolution capabilities enable a publisher to provide information about any or all of those products via that CrossRef DOI.

Diving Deeper, Getting Granular

Another important advance in EPUB 3 is the ability to associate metadata—and specifically, identifiers—below the “package” level: not just applying to the EPUB as a whole, but to components or even locations within it.

First, a few words about the metadata mechanisms in EPUB 3. In order to keep the specification as simple as possible while enabling the rich metadata that some publishers will want to include, EPUB 3.0 specifies two basic mechanisms for expressing metadata:

- The Dublin Core Metadata Element Set (DCMES),¹⁴ a basic vocabulary of fifteen properties that is widely used to describe resources in publishing, media, and libraries.
- A general purpose <meta> element that has a variety of “properties” defined in the specification. This is a simple, elegant, and extremely versatile method of incorporating diverse metadata into an EPUB without burdening the spec—and its users—with an unduly complex metadata model.

Most of the metadata in an EPUB resides in the <metadata> element, which is the first element within <package>, the root element of the Package Document (.opf) and the EPUB component whose main purpose is to provide information about the EPUB as a whole. The next element within <package>, following <metadata>, is <manifest>. The manifest is where the “Publication Resources” in the EPUB—the XHTML and other content files (e.g.,

¹³ This is quite well aligned with (though not identical to) CrossRef’s guidelines on distinguishing between a “major revision” and a “minor revision”: a major revision contains changes significant enough that they affect the interpretation of the work and are called to the reader’s attention; a minor revision does not. In both EPUB and CrossRef, and of necessity, a large degree of judgment is left up to the publisher.

¹⁴ See <http://dublincore.org/documents/dces/> for the authoritative documentation of DCMES.

chapters), the images, the videos and audio files, *every* individual component contained in the EPUB—is listed and identified, as an <item/> in the <manifest>.

Here is where there is another opportunity to use the CrossRef DOI. In the EPUB 3.0 spec, it may not be immediately obvious how to do this, because the <meta> element can only be used in <metadata>, not in <manifest> and not on <item>. This is how to do it:

- Each <item> in the <manifest> has an @id, e.g. <item id="chapter5" . . . />
- Use a <meta> element in <metadata> to “refine” that <item>, like this:
<meta refines="#chapter5" property="dc:identifier"> http://dx.doi.org/10-1234/MyBookChapter5</meta>

In plain English, that says “associate the identifier <http://dx.doi.org/10-1234/MyBookChapter5> with the <item> that has the @id “chapter5”.” (There is more in the EPUB spec that shows how to provide more explicit reference to the “identifier-type” and its @scheme, but by using the <http://dx.doi.org/> prefix as in the example, above a doi-aware system should handle it properly.)

You can even use this same strategy to point to a node *within* that document (for example, if you have assigned a CrossRef DOI to a portion of content below the document level, e.g. a <section> that is an important section on its own), like this:

- <meta refines="chapter5.html#[id of that section]" property="dc:identifier"> <http://dx.doi.org/10-1234/MyBookChapter5ImportantSection></meta>

Note that you can do this with any item—not just chapters. Since *every* individual resource you put into your EPUB is identified by an <item> in the <manifest>, if you have an image, or a video, or some other resource to which you’ve assigned a DOI, you can associate it the same way.

Incorporating ONIX, MARC, MODS, and XMP

Unlike virtually all other eBook formats, EPUB 3 enables the inclusion of a full ONIX file or MARC record as part of the EPUB. These are considered “external resources” (even though they may be within the .epub file), so they are not part of the <package> (the .opf file) that contains the metadata that is considered intrinsic to the EPUB, and they are not <item>s listed in the <manifest>. Instead, they are referenced by the <link> element in the <metadata>.¹⁵ Formally, the EPUB 3.0 spec recognizes these metadata formats and provides the following vocabulary for the values of properties in the @rel attribute on <link>:¹⁶

- *marc21xml-record*, a MARC record expressed as XML (see <http://www.loc.gov/standards/marcxml/>)
- *onix-record*, the international ONIX for Books format maintained by EDItEUR (see <http://www.editeur.org/8/ONIX/>)

¹⁵ See <http://idpf.org/epub/30/spec/epub30-publications.html#elemdef-opf-link> for the detailed specifications of how to do this.

¹⁶ See <http://idpf.org/epub/30/spec/epub30-publications.html#sec-link-rel-values>, where a fifth property, xml-signature (for digital signatures) is also included.

- *mods-record*, the Metadata Object Description Schema from the Library of Congress (see <http://www.loc.gov/standards/mods/>)
- *xmp-record*, the Extensible Metadata Platform typically used to embed metadata in graphic objects like images; it has become standardized as ISO 16684-1:2012 (see <http://www.adobe.com/products/xmp/>)

DOIs can be included in all of these.

Permanent vs. Volatile Metadata

There are pros and cons about embedding some types of metadata records or portions of their constituent metadata within the EPUB itself. Some metadata is “permanent,” always true of the EPUB (typically, bibliographic metadata of the sort mainly used by CrossRef, like author names, book or chapter titles, and identifiers, as well as most technical metadata). But some metadata is “volatile,” changing from time to time (commercial, promotional, and administrative metadata such as pricing, sales territories, usage rights, marketing information like reviews and awards, and even certain information about authors, like their titles, affiliations, contact information, and bios). And some metadata is for technical or administrative purposes and is not intended to be made public.

The most important example of a common metadata record that can contain volatile information is the updated ONIX record that publishers commonly provide to the supply chain on a periodic basis, often weekly, monthly, or quarterly. These ONIX records seem “finished” when the book is published, but they’re usually not. If they’re embedded as a whole in an EPUB, that EPUB—in fact, all copies of it that have found their way to end users—may eventually have some outdated or even incorrect metadata.

It is often overlooked that traditionally stable items like prices, sales territories, and usage restrictions are much more volatile in eBooks than they are for print. Care should be taken not to permanently embed metadata that is going to change from time to time in a way that will live forever in a user’s EPUB file. On the other hand, linking to these metadata resources over the web makes them available only when the eBook is online, and makes them less useful to the reading system (which will be unlikely to depend on them or even use them at all if they’re not part of the offline package).

4.0 Recommendations for the Use of DOIs for Scholarly Books in the Mobile Environment

The publishing landscape has never been more diverse and dynamic—which is the positive way of saying “complicated and unpredictable.” The CrossRef DOI provides both an anchor and a lighthouse in these turbulent waters. It offers publishers an unparalleled mechanism that enables users to find, cite, acquire or access, and use their content in whatever forms, on whatever devices or platforms, are available, now or in the future.

While this report has suggested a wide varieties of the ways in which the CrossRef DOI can be used for books, no publisher will find it practical or even possible to do all of these things. Here are recommendations for how scholarly publishers can use the CrossRef DOI to best advantage in today’s increasingly mobile-dominated environment.

Register Book-Level DOIs

The most basic and obvious use is to register a CrossRef DOI for each title, ideally with a Registration Agency that provides both reference linking and multiple resolution (as CrossRef does). This can be done either by depositing metadata records or by accessing CrossRef’s web deposit form,¹⁷ which requires no knowledge of XML. Once a publisher has joined CrossRef and has been assigned a DOI prefix or prefixes (which can be done through CrossRef), it can then create whatever DOIs it chooses.

When a DOI has been assigned by the publisher to a particular book, the metadata required to register that DOI for a monograph with CrossRef is:

- *Title**: The title of the book.
- *Book DOI and URL**: The book-level DOI and the URL the DOI should resolve to.
- *Contributors**: At least the lead author name, but ideally all authors or editors.
- *Publisher**: The name of the publisher as it would be most often used in a citation.
- *ISBN**: Currently required by CrossRef for web form deposits.¹⁸
- *Edition*: Just the number of the edition, if applicable.
- *Publication date**: Year is required, month and day are optional

The asterisks indicate required metadata; but best practice is to supply as much metadata as possible, because the more metadata you have associated with a book the easier it will be to match queries to it—requests to identify the CrossRef DOI for a certain book—which will enable others to find and use your CrossRef DOI, most commonly in citations.

¹⁷ See http://www.crossref.org/help/CrossRef_Help.htm and navigate to “Depositing using a web form.”

¹⁸ Note that this currently requires an ISBN, which is technically only for one product or version of that title. CrossRef used to recommend using the ISBN as a suffix when registering a DOI, and there are a great many existing DOIs “in the wild” that do so. However, as the eBook ecosystem has evolved, this practice can introduce unfortunate ambiguities. See the BISG identifier recommendations at http://www.bisg.org/docs/BISG_Policy_1101.pdf and the International ISBN Agency’s guidelines for assignment of ISBNs to eBooks at <http://www.isbn-international.org/faqs/view/17>.

Use Those DOIs Everywhere You Can

DOIs are an ideal viral identifier: they can be put virtually anywhere and still “work.” Of course they should be included in all metadata provided to retailers, aggregators, distributors, and others in the supply chain, whether in ONIX 3.0 (the ideal form) or in whatever way a partner requires it. DOIs should also be included in your catalogs, whether online or in print, and incorporated as much as possible into any eCommerce systems you use. Make sure they’re in any citations of the book (notes, references, bibliographies) in other books you publish. Finally, don’t overlook incorporating them in all marketing and publicity done for a title, including e-mails, especially those that accompany advances or review copies. This will encourage the recipients to use them—in *their* e-mails, *their* reviews, *their* blogs. We are in an era of social media. The DOI should be very infectious. You want it to spread.

Exploit Multiple Resolution

Book publishers often refrain from registering DOIs because they mistakenly believe the DOI has to take a user directly to the book’s content, and if their books aren’t online or are behind a paywall, this appears impossible. Don’t let this be an obstacle. At a minimum, you can simply provide the URL of your website when you register the DOI; thanks to the architecture of the DOI, you can always update that later—and you should, ideally directing users not just to your website but right to the page pertaining to that book.

While it may take some updating of your website or eCommerce site, making use of the CrossRef DOI’s multiple resolution feature is hugely beneficial. That’s what enables you to show, to a user coming to you via a book-level CrossRef DOI, what their options are: what formats are available, whether the book is available only in whole or in part, as part of a subscription scheme, etc. It can also point to multiple places where the book can be obtained: your own eCommerce site, Amazon, the iBookstore, B&N.com, etc. It can even provide access to things like the author bio, sample content, other related books, etc.

Register Chapters

Especially for scholarly publishers, registering CrossRef DOIs for the individual chapters often makes a lot of sense as well. While this is most obviously important for edited collections and multi-author books with individual chapter authors, it can be good for almost any type of scholarly book. Registering chapter-level CrossRef DOIs is beneficial in three ways:

- Making it possible to cite a chapter, rather than just the book as a whole. This will be much more desirable for scholars, students, and readers.
- Making it possible to sell or provide access to chapters individually. Especially when chapters are focused on distinct topics, as they often are, a user may want just a single chapter, or select chapters, rather than the whole book. Where they might not want to buy the book, they might pay for access to the chapters they need. And “short form” eBooks are becoming ever more popular.
- Turning a chapter or a selection of chapters—perhaps even from several different books—into a new publication, offered as an eBook, can become a very compelling

proposition. The resulting derivative book can be useful to readers who wouldn't buy the individual books the chapters are taken from. This is an excellent way to get more mileage out of the investment in publishing the books in the first place. The eBook environment provides a low-cost, low-risk way to do this. When each chapter has its own DOI, the administrative process for such derivative products is greatly streamlined. (One example: royalty tracking and management.) In addition, the CrossRef DOI associated with each chapter can direct a reader back to the book it was originally a part of, not to mention other books or publications it may relate to—a big benefit both to the reader and to the publisher.

The CrossRef Help page cited above provides information on how to register CrossRef DOIs for book chapters. The metadata required is minimal, especially if the book is already registered.

What About Getting Even More Granular?

Registering CrossRef DOIs for books at a level below the level of the chapter is not usually an across-the-board decision. It's best to look at what component parts within a book might have a viable life outside the book—for example, a compelling image or a particularly instructive diagram useful to scholars or students independent of the book for which it was created. Keep in mind, though, that there are rights issues associated with these kinds of components. If you licensed them, you probably don't have the right to distribute them, but if you created them and own the rights, then *you* might potentially be able to license them. If you have the demand and the systems to support this, then the CrossRef DOI is extremely valuable.

Put the DOIs in the eBooks

The book-level CrossRef DOI is the ideal identifier to use in metadata *about* your books. But what about the CrossRef DOI going *in* an eBook?

At a minimum—if you have registered one and only one CrossRef DOI for a book—then using it as the identifier in the various eBooks you produce is logical. This addresses the continuity issue, enabling all the different versions and formats of the book to share that one common identifier. Because those versions and formats should each have a unique ISBN, the book-level CrossRef DOI provides the common identity that ties them together.

Note that this in no way precludes other identifiers from being included in the eBook as well. The new EPUB 3 spec is particularly useful in this regard. It accommodates any number of identifiers in any scheme. While at present most reading systems don't have good mechanisms to make use of those multiple identifiers, they do no harm, and make the EPUB file more useful to the various systems in which it may be incorporated over time. Using the CrossRef DOI as the primary identifier for all e-books, especially in the form of a URL (e.g., <http://dx.doi.org/10.1234/xyx56789abc>), makes it much more likely that it will be accessible and useful to the distributor, the library, the user—and the publisher.

Use the DOI in Citations

For scholarly publishers, the single most important use of the DOI is for reference linking. Those who publish journals do not need to be told this.

It is time to put them in the citations in books.

Not just the DOIs of your own books. Not just DOIs of books in general, or chapters. Any and all available DOIs—for books, for chapters, for journal articles, for datasets, for conference proceedings, for reports. Any DOI that can be obtained from CrossRef and other RAs should be incorporated into every possible citation *in your books*.

Book publishers now recognize the value of their digital files. But they persist in thinking of their books as existing in a print-centric universe. Although eBooks have become commonplace, they are seen as existing as self-contained objects that don't interact dynamically with the scholarly ecosystem in the way journal articles do. And there is the misconception that there simply aren't enough DOIs to link to.

This is all far from the truth. There is a torrent of DOIs being issued for things cited in scholarly books—CrossRef currently manages over 50 million CrossRef DOIs and registers over half a million new ones every month, and there are already well over 4 million CrossRef DOIs for books and book components, especially chapters. These are all available for incorporation into citations *right now*.

CrossRef provides good documentation of a number of ways to incorporate DOIs in reference lists—ranging from using the actual URL form of the DOI to using graphics or text or both with the DOI behind them—at http://www.crossref.org/02publishers/doi_display_guidelines.html.

And because of the power of the DOI, any citation containing a CrossRef DOI—in a footnote, in a reference list, in a bibliography, whether in a print book, an online book, or an eBook—is *actionable*: a user can be directed to the cited resource from all of them.

CrossRef DOIs not only increase the *visibility* of books in the scholarly ecosystem, they can also improve their *status*. Increasingly, citation counts and citation metrics rely on the DOIs. Publications lacking DOIs can be ignored by the systems that generate these metrics, which can lead to a significant underestimation of the impact of your books.

We need to make DOIs as indispensable to scholarly books as they are to journals. We need the scholarly ecosystem to be so digitally interconnected that scholarly books that lack DOIs seem as invisible as journal articles that lack DOIs. The time to do this is now.

5.0 Resources

The following resources provide useful documentation of the identifiers and standards discussed in this report, as well as more extensive discussion of the issues associated with them and recommendations for their use.

Identifiers

GENERAL

The “Roadmap of Identifiers” published and periodically updated by the Book Industry Study Group (BISG), available at http://www.bisg.org/docs/Roadmap_of_Identifiers.pdf.

For a useful discussion of issues surrounding identifiers for works and components of works, see the June 2011 discussion paper “Requirements for the Identification of ‘Components’” by Mark Bide and Graham Bell of EDItEUR, available at <http://www.editeur.org/117/Publications/>.

For those interested in the differences between the two leading abstract models underlying most bibliographic metadata schemes, the <indec> metadata framework is discussed in detail at http://www.doi.org/topics/indec/indec_framework_2000.pdf and FRBR, Functional Requirements for Bibliographic Records, is discussed in detail at <http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records>.

DOI

The authoritative source of information about the DOI is the International DOI Foundation (IDF), <http://www.doi.org>. There you will find a rich assortment of resources about the DOI, the various DOI registration agencies, examples of DOIs for various types of content, and helpful documentation, white papers, factsheets, and FAQs. Most importantly, the IDF’s *DOI Handbook*, <http://www.doi.org/hb.html>, is the essential source for the most comprehensive information about the DOI.

The publication of the DOI as ISO 26324 is available at http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43506.

NISO’s “ANSI/NISO Z39.84-2005, *Syntax for the Digital Object Identifier*, is available at http://www.niso.org/kst/reports/standards?step=2&gid=None&project_key%3Austring%3Aiso-8859-1=a83b89410e67c2d7f6c7f43a6183d3f18e8c3195.

For more information on the shortDOI, go to <http://shortdoi.org/>.

ISBN

The authoritative source for information about the ISBN, the International Standard Book Number, is the International ISBN Agency, www.isbn-international.org. Like the IDF for the DOI, the International ISBN Agency provides a wealth of information about the ISBN. Of particular note is the *ISBN Users’ Manual* at <http://www.isbn-international.org/page/info>.

For use of the ISBN in the United States, see the Bowker site, www.isbn.org.

For use of the ISBN in the UK, see the Nielsen site, www.isbn.nielsenbook.co.uk.

Information about other national ISBN agencies is available at <http://www.isbn-international.org/agency>.

For information about ISBN-A, see “The ISBN System in Relation to the DOI® System,” a DOI System Factsheet available at <http://www.doi.org/factsheets/ISBN-A.html>.

ISTC

The authoritative source for information about the ISTC, the International Standard Text Code, is the International ISTC Agency, <http://www.istc-international.org>.

Information about the various ISTC Registration Agencies is available at <http://www.istc-international.org/html/registration.aspx>.

Metadata Standards

DUBLIN CORE

The authoritative source of information about Dublin Core metadata is the Dublin Core Metadata Initiative (DCMI), <http://dublincore.org>.

ONIX

The recommended standard for book supply chain metadata is ONIX for Books 3.0, available at <http://www.editeur.org/12/About-Release-3.0/>. This site provides extensive documentation, code lists, and a wealth of other useful information about ONIX.

MARC

Extensive information about the U.S. Library of Congress’s MARC (MACHINE Readable Coding), including MARC21 and MARCXML and MARC’s relation to other standards, is available at <http://www.loc.gov/marc/>.

MODS

The Metadata Object Description Schema (MODS) is also available from the Library of Congress, at <http://www.loc.gov/standards/mods/>.

METS

Another very useful general purpose metadata framework, the Metadata Encoding and Transmission Standard (METS), is also available from the Library of Congress, at <http://www.loc.gov/standards/mets/>.

XMP

For information on the eXtensible Metadata Platform, XMP, see <http://www.adobe.com/products/xmp/overview.html>.

METADATA IN EPUB

A useful comprehensive overview of the EPUB 3.0 standard is at <http://idpf.org/epub/30/spec/epub30-overview.html>. The information about metadata in EPUB 3.0 is mostly covered at <http://idpf.org/epub/30/spec/epub30-publications.html>.

CrossRef

General information about CrossRef is at <http://www.crossref.org/>. There you will find extensive resources and tools relating to the use of the DOI in general and in the context of CrossRef's services as a DOI Registration Agency.

Of special note are the excellent resources available at http://www.crossref.org/help/CrossRef_Help.htm. This is the best single resource regarding the DOI for scholarly publishers.

CrossRef's recommendation of best practices for the use of DOIs in books is available at http://www.crossref.org/06members/best_practices_for_books.html.

Guidelines for the display of the DOI are at http://www.crossref.org/02publishers/doi_display_guidelines.html.

A helpful webinar, "CrossRef DOIs for eBooks," is available at <http://www.slideshare.net/CrossRef/crossref-dois-for-books-or-how-readers-can-find-your-stuff>.

DRAFT