

# DIGITAL OBJECT IDENTIFIERS FOR EBOOKS: WHAT ARE WE IDENTIFYING?

**W3C Position Paper<sup>1</sup>**  
**Stephen Mooney<sup>2</sup>**  
**17 December 2000**

## **Background: DOI-EB**

Developed by the International DOI Foundation (IDF), the Digital Object Identifier (DOI) is an identification system for intellectual property in the digital environment.<sup>3</sup> The IDF recently established a working group entitled *Digital Object Identifier for e-Books (DOI-EB)*, the purpose of which is to:

- determine requirements with respect to the application of unique identifiers to eBooks
- develop proofs-of-concept for the use of DOIs with eBooks
- develop technical demonstrations, possibly including a prototype eBook registration agency.

Since commercial implementations of DOI will be an important focus of DOI-EB, the active participation of all parties in the value chain is welcomed.

## **What is an eBook?**

Before seeking to apply unique identifiers to eBooks, one might first ask, “What exactly is an eBook?” There are no covers, as it were, and the very attributes of an eBook may make defining the term more difficult and less precise. That no well-developed eBook market exists no doubt

---

<sup>1</sup> Portions of this document are derived from a presentation given by Steve Mooney to the International Rights Directors’ Meeting at the 2000 Frankfurt Book Fair. Permission was granted to the Frankfurt Book Fair to copy and distribute the presentation as originally given. A great debt is owed to Eamonn Neylon for his comments on that presentation. The initial draft of this paper was prepared in November 2000 and was not published.

<sup>2</sup> Pye Brook 47, Boxford, Massachusetts 01921, USA, +1 978 561 1036, stephenmooney@yahoo.com.

<sup>3</sup> See Bide, Mark, *In Search of the Unicorn, The Digital Object Identifier from a User Perspective*, BNBRF Report 89, Book Industry Communications, London, February 1998. See also Bernstein, Paula, *DOI: A New Identifier for Digital Content*. Searcher: The Magazine for Database Professionals, Vol. 6, No. 1, Jan. 1998. See also *The DOI Handbook*, version 0.5.1, The International DOI Foundation, Washington, D.C. USA and Geneva, Switzerland, September 2000.

adds to the difficulty in defining what we mean by “eBook.” The term “eBook” may well turn out primarily to be a marketing term, one that proves to be quite inadequate in the long run because, ironically enough, the greatest potential of eBooks is not to be books at all.<sup>4</sup>

In spite of the obstacles, brave souls have attempted, not completely successfully as yet, to define the term “eBook.” Lest one conclude that I cast stones from a glass house, let me make clear that I have been closely associated with two of these efforts.<sup>5</sup> The draft definition of the Electronic Book Exchange Working Group (EBX)<sup>6</sup> indicates that an eBook may consist of a single page, and it also refers to paper books to define electronic books. It needs work. A somewhat different approach is taken by the Open eBook Forum (OeBF) that provides a draft definition<sup>7</sup> that is appealing because it refers to unique identifiers and metadata, although that may not be strictly necessary. Further, neither it nor the EBX definition makes a distinction between an abstraction of a work and a particular manifestation of that work.<sup>8</sup> Still a third definition is offered by ONIX which takes a more subtle approach by describing three ebook entities.<sup>9</sup> Interestingly, and helpfully, the ONIX approach separates content from presentation.

---

<sup>4</sup> At the American Association of Publishers Professional and Scholarly Publishing (PSP) meeting in February 2000, Keiron Hylton, Managing Director at Berkery Noyes & Co. spoke of books (and journals incidentally) as no longer being free standing units of knowledge but as becoming mere points of access – much like a web browser. Hylton speculated that such would be the case within ten years time, and his views – which I find to be compelling – pose interesting challenges for those committed to thinking about eBooks as mere electronic analogs of paper books.

<sup>5</sup> I am Vice Chair of the Electronic Book Exchange (EBX) Working Group and participate actively in the Open eBook Forum, most recently serving in the Digital Rights Management (DRM) Working Group and as Leader of the Identifier Special Interest Group.

<sup>6</sup> The EBX draft defines an eBook to be “...a digital object that is an electronic representation of a book. While an eBook can consist of a single page, it is normally thought of as an electronic analog of a multi-page hardcover or paperback book. An eBook may exist in a variety of formats...” *The Electronic Book Exchange System*, version 0.8, Book Industry Study Group, New York, New York, USA, July 2000, at 6, [www.ebxwg.org](http://www.ebxwg.org)

<sup>7</sup> The OeBF draft defines an e-Book as “...short for *electronic book*. 1. A Literary Work in the form of a Digital Object, consisting of one or more standard Unique Identifiers, metadata, and a Monographic body of content, intended to be published and accessed electronically. 2. May also refer to the hardware device created for the purpose of reading eBooks (RocketBook, SoftBook, Franklin e-bookman),” *A Framework for ePublishing Ecology, Public Comment Draft, version 0.78*, The Open eBook Forum, Boulder, Colorado, USA, 25 September 2000, at 6, [www.openeBook.org](http://www.openeBook.org)

<sup>8</sup> First commented upon by Norman Paskin, Director, International DOI Foundation, Washington, D.C. USA and Geneva, Switzerland and Mark Bide, Mark Bide & Associates, London, United Kingdom.

<sup>9</sup> The ONIX draft speaks of:

- “...EBook content – a package of text and other content which a publisher puts together in order to be realized in one or more eBook formats...”
- “...EBook rendering – a realization of an eBook content package in a particular format for a particular eBook reader or family of readers, whether for downloading for local use of by remote online access...”

In grappling with the term “eBook” we must also note that the focus of ebook transactions, for the time being, will be on the initial sale and rights associated with the initial sale.<sup>10</sup> The term “eBook” serves an immediate marketing need in that it handily identifies an opportunity many seem to understand. Yet, in the long run, the very term “eBook” may be a misnomer. As the marketing focus shifts from the initial sale of an ebook to secondary uses and licenses involving more detailed rights transactions, it is very possible that term “eBook” may be much less useful than it appears to be at the moment. At that point, we may need to speak of something other than eBooks, and only by means of unique identifiers and associated metadata can these rights transactions come to fruition in the market.<sup>11</sup>

Merely selling eBooks to consumers does in fact *not* require persistent, unique, and global intellectual property identifiers such as DOI. Indeed, Barnes & Noble, Glassbook/Adobe, netLibrary, Microsoft, Digital Goods and others sell eBooks on the web every day without such identifiers.<sup>12</sup> However, few observers particularly expect current eBook sales channels to generate significant profit in the near term, and many, of whom I am one, suggest that selling or licensing eBooks *profitably* may indeed require persistent, unique, and global identifiers. Accordingly, we must focus not simply on discovery of eBooks on the Internet but also on the exploration and exploitation of profitable services that persistent, unique, and global identification of intellectual property will facilitate along the entire eBook value chain. Such a focus causes us to ask a second fundamental question, “What does it mean to sell an eBook?”

- 
- “...EBook component – a portion of text or other content which is used or available to be used in electronic form as part of an “eBook content package...”

Draft *ONIX International eBook Proposals for Release 1.2*, 22 September 2000, Editeur c/o Book Industry Communication, London, United Kingdom, 22 September 2000 at 3.

<sup>10</sup> This is not to disregard the many efforts currently under way that focus on more granular rights transactions. However, such efforts have captured neither imagination nor investment to the extent that eBook have.

<sup>11</sup> See Paskin, Norman. *Digital Object Identifier: Implementing a Standard Digital Identifier as a key to effective Digital Rights Management.*, International DOI Foundation, Washington, D.C. USA and Geneva, Switzerland, April 2000. See also Erickson, John. *The Role of Metadata Supply Chains in DOI-Based, Value-added Services.* ICSTI Forum, Quarterly Newsletter of the International Council for Scientific and Technical Information, No. 30, April 1999. See also Bide, Mark and Rust, Godfrey, Interoperability of data in e-Commerce Systems, [www.indecs.org](http://www.indecs.org)

<sup>12</sup> Microsoft has announced that it will support DOI in its eBook reader.

## What does it mean to sell an eBook?<sup>13</sup>

What one considers to be “selling” an eBook may, in certain cases, be licensing. I make the point here not for the purpose of legal analysis but to explore commercial possibilities that unique identifiers for eBooks could facilitate. Mere discovery of intellectual property entities does not inform the commercial transaction one intends. Thus, we should consider what commercial models or options should follow discovery. At least three factors present themselves in exploring whether an eBook transaction is a sale or a license: encryption technology, migration, and breaking an eBook into its parts.

The first factor is encryption technology. Encryption technology generally comes part and parcel to the ebook. When one purchases a paper book, there is a notion of free transferability. That is to say, I can give the paper book to another or I may loan it. To the extent that encryption technology restricts the free transferability of the ebook, it has features that more closely resemble a license and not a sale. I am not suggesting that this must be so in every case, merely that this must be considered in planning ebook businesses.

The second factor is migration.<sup>14</sup> What happens when the format of my existing ebook becomes obsolete or a particular device manufacturer goes out of business and a format is no longer supported? Do I have the right to migrate the work to a successor format? Or must I purchase a new version? A sale is characterized by continuing access, as is the case with paper. If my access is not continuing, if I cannot migrate the work, the transaction takes on the characteristics of a license.

The final factor is breaking an eBook into its “parts.” With a paper book, if I consider the whole thing rubbish with the exception of Chapter 6, I can remove Chapter 6 from the book, put it in my filing cabinet, and throw the rest away. May I do so with an eBook? May I put it in my “e-

---

<sup>13</sup> See my comments in *Minutes of the Technical Committee of the Electronic Book Exchange (EBX) Working Group*, 11-12 July 2000, Maynard, Massachusetts, USA. See also my three additional presentations: *DRM and Digital Distribution for Publishing*, San Francisco, August 2000, *International Rights Directors Meeting*, Frankfurt, October 2000, and *Forum on DOI*, Frankfurt, October 2000. I would be happy to supply these upon request.

<sup>14</sup> See *Preserving Digital Information, Report of the Taskforce on Archiving of Digital Information*, Commission on Preservation and Access and the Research Libraries Group, Mountain View, California, USA, 1 May 1996.

filing” cabinet? If not, the transaction takes on the characteristics of a license, and not a sale. Might this also inform what eBook “parts” need to be uniquely identified?

### **Issues for DOI-EB**

It may not be necessary to detail every last branch of the tree to gain an understanding of the tree, as Mark Bide pointed out recently by way of analogy.<sup>15</sup> Merely scoping out the trunk and major limbs will facilitate a greater understanding of the tree. Among the issues DOI-EB must explore are:

- Why would one use an identifier? As noted above, eBooks are being sold today without identifiers such as DOI, so they may not be strictly necessary at one level. Still, the decision of Microsoft to support DOI in its eBook reader leads one to conclude that commercial enterprises appear to see value in the approach DOI-EB will explore. An understanding of why Microsoft, as an example, decided to use an identifier needs to be achieved.
- What entities within the entire eBook value chain must be identified? Is it sufficient simply to identify the eBook itself and be done with it? One participant in the value chain with whom we spoke<sup>16</sup> understands article level application of DOIs with journals, but with books he “sees the need for going down to the graphic level” to identify components of an eBook.<sup>17</sup>
- How will DOIs fit into eBook workflow? What are the components? “DRM begins upon creation and production workflow,” according to an individual with whom we spoke. How DOIs need to get embedded in commercial workflows is of key interest.
- What syntax will address fragments of works down to the graphic, table, and supplementary materials level of eBooks?
- What metadata is mandatory?<sup>18</sup> One publisher has suggested the virtue of clicking on a DOI so that one may be brought to a page that shows: where the DOI was assigned in the work process, who owns what rights, and what formats are available

---

<sup>15</sup> In conversation, October 2000.

<sup>16</sup> Eamonn Neylon and I conducted a series of initial discussions about DOI-EB with publishers and others in the value chain.

<sup>17</sup> Such an approach would be essential for a publisher who presented me outside the context of DOI-EB with the following issue. The publisher wanted to enable a professor to select from among chapters of their own books, as well as books of other publishers, to create custom eBooks online. The widespread commercialization of such an endeavor absolutely requires the application of metadata and unique identifiers to the underlying intellectual property at the necessary levels.

<sup>18</sup> See ONIX International, Editeur, London, United Kingdom, [www.editeur.org](http://www.editeur.org). See also Lagoze, Carl. *Accommodating Simplicity and Complexity in Metadata: Lessons from the Dublin Core Experience*, Presented at

- May one sell pieces of an ebook? What are the “saleable” pieces? How are they identified?
- One participant in the value chain noted that there is actually no way to tell that an eBook published in the United Kingdom and one published in the United States are in fact the same eBook. There is no global identifier at present.
- What is the customary bundle of rights that should go along with eBook transactions? Consumers need to have a common expectation of what they are purchasing when they buy an eBook. This goal is not impossible, but exploration of it very much implicates the types of services players in the eBook value chain will offer, or even contemplate.

Upon reviewing eBook wares shown at the 2000 Frankfurt Book Fair, one observer claimed that, “...the international publishing industry is coping with a severe case of vertigo...”<sup>19</sup> Much dizziness no doubt is due to the wealth of opportunity before us. That no single vendor solution has yet captured the attention of so young an industry is not particularly surprising. That so many solutions focus on “electronic analogs of paper books”, however, is. Unique, persistent, and global identification of eBooks will surely facilitate the greatest potential that eBooks have - not to be books at all.

---

Seminar on Metadata, Organized by Archiefschool, Netherlands Institute for Archival Education and Research, June 8, 2000. See also Neylon, Eamonn., *EBX Metadata Specification*, version 0.4, 23 October 2000, draft currently under review by EBX Technical Committee.

<sup>19</sup> Felici, Jim, *E-books on the Ramparts at Frankfurt Book Fair 2000*, The Seybold Report on Internet Publishing, November 2000, vol. 5 no. 3, Seybold, Media, Pennsylvania, USA, p. 29.

A digital object identifier (DOI) is a serial code used to uniquely identify content of various types on electronic networks. The DOI system is particularly used for electronic documents such as journal articles. The DOI system began in 2000.[1]. Metadata about the object is stored in association with the DOI name. The prefix identifies the registrant of the name, and the suffix is chosen by the registrant and identifies the specific object associated with that DOI. Most legal Unicode characters are allowed in these strings, which are interpreted in a case-insensitive manner. For example, in the DOI name 10.1000/182, the prefix is 10.1000 and the suffix is 182.