



Patricia Seybold Group

Trusted Advisors to Customer-Centric Executives

Protecting Your Digital Assets

Technical Journal Publishers Lead the Way
Using Digital Object Identifiers

By Patricia B. Seybold



Protecting Your Digital Assets

Technical Journal Publishers Lead the Way Using Digital Object Identifiers

By Patricia B. Seybold

February 13, 2003

NETTING IT OUT

The publishers of scientific, medical, and technology journals have long been leaders in the push for standards to uniquely identify digital assets. One such standard, which is being used extensively in journal publishing, is the Digital Object Identifier (DOI). The DOI is a standard mechanism to provide a unique identifier to granular digital assets, such as chapters within a book, single tracks of music on a CD, and individual images. Currently, the DOI standard is in use by over 170 journal publishers worldwide, and over 6 million articles across 6,800 journals are being tagged with DOIs.

We predict that, within five years, the DOI standard will be used to tag any “published” material from any industry—that is, all content or information that is officially released for consumption, whether within or outside of your firewalls.

IDENTIFYING & TAGGING DIGITAL ASSETS

While the music, software, commercial publishing, and the movie industries all agonize about how to prevent unauthorized copying and distribution of their copyrighted materials, technical journal publishers have made great strides in managing, licensing, distributing, and promoting their digital assets. Last week, we discussed the new business models that the publishers of scientific, medical, and technical journal have put into place to take advantage of the Internet and emerging customer needs and expectations¹. The good news is that the infrastructures

¹ See “Understanding Digitization: Trends in Business Models,”

and the business models that journal publishers and research librarians have put in place can be easily adopted by other ancillary industries.

The bad news is that these industry segments are so disparate, serving such different audiences (academic researchers vs. mass consumer), that cross-fertilization is bound to be difficult.

Using Digital Object Identifiers (DOIs)

In order to protect intellectual assets, they must be easy-to-identify. Each creation must be uniquely identified, registered, and adequately described so that it can be easily differentiated.

The scientific and technical publishing community was among the leaders in the electronic publishing world to push for standards to uniquely identify each digital asset.

Publishers had always needed a way to uniquely identify (and to protect) their intellectual assets. For generations, the Library of Congress in the U.S. filled the role of uniquely cataloging each published work, using ISBN (International Standard Book Number) numbers for books, and ISSN (International Standard Serial Number) numbers for serials and periodicals, among others. But these identifiers are not “actionable” – you can’t click on them on the Net and do something with them. And what about the individual articles that make up a journal? Or the chapters in a book? Or the tracks of music on a CD? Or a particular performance of Shakespeare’s *Twelfth Night*? Identifying any object, in a persistent actionable way, is the role of the DOI: Digital Object Identifier.

<http://www.psgroup.com/doc/products/2003/2/BP2-6-03CC/BP2-6-03CC.asp>

The need for unique object identifiers became evident very early in the history of electronic publishing. Once material began to be digitized, it quickly became apparent that a more granular classification scheme would be required.

THE EVOLUTION OF THE DOI. In the mid-'90s, several organizations, including the American Association of Publishers, the International Publishers' Association, and the International Association for Scientific, Technical, and Medical Publishers, began working jointly on the development of a standard mechanism to identify uniquely granular digital assets (e.g., articles, tracks of music, images, etc.). Not only did these groups agree on a standard, they also tackled the problem of ensuring that it would be implemented. They modeled their process on the W3C and the Bar Code development and implementation process. In 1998, they formed a foundation—the International Digital Object Identifier (DOI) Foundation—to provide an implementation mechanism, a set of social structures, and an educational body to ensure that the DOI standard is widely implemented. DOI is an implementation of the concept of a Uniform Resource Identifier (URI) or Uniform Resource Name – great concepts, but ones which require practical implementations like DOI to be useful. It's also built on extensible principles, and well-proven standard-based technology derived from community efforts on interoperability of metadata, including recent ISO MPEG work.

WHAT'S A DOI? A DOI is a digital object identifier for any object of intellectual property. According to the <http://www.doi.org> Web site, "DOIs have been called 'the bar code for intellectual property.' A DOI provides a means of persistently identifying a

piece of intellectual property on a digital network and associating it with related current data...A DOI is associated with defined services and is immediately 'actionable' on a network."

Again, from DOI.org: "A name (or unique identifier) for a digital object enables that name to be resolved to one (or many) of several different pieces of data which may be associated with the digital object. Such pieces of data can be locations of the object, or services about the object, or any other defined piece of data. Resolution enables a single name (the identifier, DOI) to be used persistently to manage the object, even if any of those pieces of data (like location) change. Resolution therefore (a) enables persistence and (b) enables multiple services to be directly associated with the DOI."

There have been several standards efforts that have taken root within the electronic journal publishing community. The two that appear to be the most vibrant at this point are the Digital Object Identifier (DOI), which is used to uniquely identify each information object and make it actionable, and the Open URL, which transports context-specific metadata along with each identifier making it actionable in a local context (resolving to an institution's local cache, for example). Today's online journal publishers and journal portal providers tend to use both, since they are complementary. Over 6 million articles across 6,800 journals are currently being tagged with DOIs.

WHERE ARE THEY STORED?

DOIs can be persistently stored anywhere (online or offline). And they can be moved around. Each DOI is a unique number assigned to that object. The DOI does not necessarily correspond to the object's current physical location online (a URL, which can often be volatile and thus is not persistent). Instead, it's a pointer that's resolved at runtime using the "Handle System". In other words, the DOI can be invisibly mapped to a new current location (in fact, to many such locations), to a local cached copy of the work (using an OpenURL), or to a group of objects—such as a PDF file, a MS-Word file and an html file—each of which is an instantiation of the article represented by the DOI, or could be information about the article. The process of locating the specific instance of a DOI is called "resolution." Each DOI link is resolved at runtime, each time the DOI is referenced.

HOW ARE DOIs USED?

The technical journal publishing community is converg-

ing on the use of DOIs as the mechanism for uniquely identifying each article in a journal. One application of this DOI standard, commencing in 2000, is now in widespread use by over 170 journal publishers worldwide through a DOI implementation and agency called the Crossref system (www.crossref.org). The Foundation expects many other such applications to emerge in the near future in other sectors, from its various existing and future Registration Agencies

Over 6 million articles across 6,800 journals are currently tagged with DOIs, and it is anticipated that an additional 2 - 3 million articles will be added each year. As of December 2002, at least 2 million DOI-tagged journal articles per month were being accessed and resolved.

Associating Metadata with DOIs through Application Profiles

Each genre of DOI—scholarly articles, book chapters, paintings, digitally-recorded performances of songs—will eventually have one or more distinct Application Profile associated with it. Each Application Profile represents both a genre of works, e.g., journal articles vs. recorded music, and the intended usage or application of the DOI. An Application Profile consists of a core set of metadata (six elements), an additional set of structured metadata elements, plus some rules (policy, business and procedural rules, not all necessarily automated).

The DOI Foundation has put enormous effort into ensuring that Application Profile metadata can be interoperable on a structured basis, using some of the recent work in ontologies now emerging. DOI metadata is not yet another metadata schema, but a way of declaring and using existing metadata schemes interoperably. This area of work is at the

heart of the Foundation's current development efforts and promises to catapult DOIs into extremely effective tools for not only identifying but managing content.

So, for example, in the scholarly publishing arena, there are at least two initiatives underway that rely on a common set of metadata that could be associated with the scholarly journal application profile: one is used to create a Web of cross-references and citations (CrossRef, an existing DOI application). The second is used to monitor usage (Counter). We'll take a look at Counter first, and then come back to CrossRef and its complementary tool, OpenURL.

Implications for Other "Publishers"

But first, what are the implications of this DOI standard for other publishers and creators, outside of the realm of scholarly journals. We think that it is profound. We predict that within five years, every article, track of music, or other digital asset will be tagged with a unique Digital Object Identifier. These will be associated with Application Profiles for each genre of digital object: photographs, movie clips, sound bytes, and so on.

If this will be true for all "published" material, what are the implications for all the content within your organization? What information assets will need to be uniquely identified? And which will escape the need to have DOIs? We suggest that you use the notion of "publication" as the criteria. Once information or content of any type is officially released for consumption, either inside or outside your firewalls, it should be uniquely identified.