9-1-2007

Digital Object Identifiers and Their Use in Libraries

Jue Wang
Western Kentucky University, jue.wang@wku.edu

The Digital Object Identifiers and the Use in the Library

Introduction

The DOI initiative was launched in October 1997 in order to develop a common mechanism to enable intellectual content management to be integrated with Internet technology at the Frankfurt Book Fair. According to the International DOI Foundation (IDF), an organization established to develop and manage the DOI System, "the Digital Object Identifier (DOI) is a system for identifying content objects in the digital environment. DOIs are names assigned to any entity for use on digital networks. They are used to provide current information, including where they (or information about them) can be found on the Internet. Information about a digital object may change over time, including where to find it, but its DOI will not change." [1] Like the Universal Product Code (UPS) bar code used on virtually every physical product, digital objects are identified by the DOIs. Since its introduction, the DOI has provided a stable, persistent link between content and a directory on the Internet to which the content owner wishes the DOI to point instead of a Web address or URL.

The DOI is also used as an important emerging international standard for identification of published material online. Today, many scientific journal publishers are using the DOI to enable readers to move seamlessly across the Internet from one location to another. Using DOI makes managing digital objects in a networked environment much easier and more convenient for both publishers and their customers. The usages of the DOIs are also very broad. Even though applications of the DOI were first applied within the publishing

industry, it will go well beyond the publishing industry, especially for electronic commerce applications.

The Structure of DOI

A DOI is an alphanumeric string or name that identifies digital content, such as an e-book, a journal article or a piece of music. The DOI contains two components: the prefix and the suffix separated by a forward slash ("/"). The prefix, also known as the "Publisher ID", is assigned by a DOI registration agency (the Directory Manager) to the publisher. Organizations may choose to have multiple prefixed for each of their imprints or product lines. All prefixes begin with "10" (this distinguishes a DOI from any other implementation of the Handle System, see http://www.handle.net), followed by a period (".") and then a number designating the organization or publisher. For example, Oxford University Press is 10.1093; Elsevier is 10.1016; and IEEE is 10.1109. The suffix, also known as the "Item ID", is assigned by the publisher and can be made up of any alphanumeric sequence of characters as long as each object can be uniquely identified. The length of the number can reach 128 characters. An existing standard identification system number, such as ISBN or ISSN, may be incorporated in a DOI by using as the suffix. Here is an example of incorporating ISBN into DOI: the ISBN is 0071381317 for the electronic version of the book "21 Leaders for the 21st Century -- How Innovative Leaders Manage in the Digital Age" by Fons Trompenaars and Charles Hampden-Turner published by McGraw-Hill and the DOI for this eBook is 10.1036/0071381317 (10.1036 represents McGraw-Hill).

The DOI can also be applied at any level of "granularity" or any file types (such as text, image, or audio-video). For a publisher, it can be applied to a whole book and also to every chapter, every illustration, photograph or table. In the case of music, it can identify a CD collection, as well as every individual track. As an example, the DOI for Mona Lisa (painting) by Leonardo da Vinci is 10.1219/10223954 (10.1219: DOI for Corbis). The suffix directory database (the DOI repository) is maintained by the publisher. Once a DOI is registered and assigned, it will not be changed during its lifetime even if the ownership and location of an object change. This makes DOIs permanent and persistent identifiers. When a digital object has been requested, a query will be sent to the DOI server (a piece of software also called a link resolver). The DOI server will find the record of the DOI and the address of its associated object and send the location back to the user's browser and then show it on the user's screen.

DOI can be part of OpenURL to solve the problem of moving users from citations to the full-text electronic journals that may available to tem via dynamically created links. By using this new technology, patrons will be able to access the full-text journal articles or other resources available in the library with just a few clicks. Seamless access to a library's e-journal collections through this linking service has tremendously improved library service and the use of library resources.

The International DOI Foundation (IDF)

The authority administrating the DOI system is the not-for-profit International DOI Foundation (IDF). The Foundation was set up by the Association of American Publishers, which joined forces with the International Publishers Association and the International

Association of Scientific Technical and Medical Publishers. The International DOI Foundation (IDF) supports the needs of the intellectual property community in the digital environment by developing and promoting of the DOI System as a common infrastructure for content management. The Foundation will govern the use of DOIs and the IDF is governed by its members, through an elected board. Members of the Board represent a wide cross section of organizations interested in the management of intellectual property in the network environment. The Board is responsible for all aspects of management of the DOI System, particularly policy formulation and standards maintenance. Board members include Elsevier, John Wiley & Sons, Springer SBM, and etc. For updated information on IDF and DOI system management, please visit the DOI website at http://www.doi.org/. IDF administrative and metadata systems are being put into place under the guidance of the Foundation as part of an ongoing development.[2]

The Central DOI Directory – Maintained by DOI Registration Agencies (RA)

The DOI directory serves as the middleman between user and the information content holder/publisher. Since digital content may change the location or ownership over the time, a central DOI depository directory has been used for DOI system maintained by DOI registration agencies. When DOI system was first launched, the IDF was the only registration agency. Due to the expansion and development of the DOI names, more registration agencies (RA) have being franchised. The RAs should be able to implement mechanisms for quality control of DOIs and metadata registration and would have the abilities to support and promote multiple resolutions. RA will assign prefixes to new

registrants in accordance with IDF standard terms. RA will also make sure that DOIs under this prefix are loaded with corresponding URLs into a globally available resolution system nominated by the IDF. RA ensures that appropriate minimal supporting metadata for each DOI and usually charges the clients (registrants) based on prefix allocation, numbers of DOIs allocated, and/or numbers of DOIs resolved.

Each RA has a geographical basis and each RA will be required to become a member of the IDF under the Registration Agency category of membership. The IDF Board will be responsible for considering all applications submitted by candidate Registration Agencies. Once becoming a RA, it allows participation in the Registration Agency Working Group and access to all IDF materials, working with IDF members in supporting and developing the system. The current registration agencies include CrossRef, Copyright Agency Limited, *m*EDRA, Nielsen BookData, OPOCE (Office of Publications of the European Union), R. R. Bowker, and TIB (German National Library of Science and Technology). More information on application profile and area of coverage of the agencies can be found at http://www.doi.org/registration_agencies.html.

The Content Database – Maintained by the Publishers

The publisher maintains a database that contains the actual content of the DOI, which was designed to control the distribution of its information. The information maintained on the publisher side may actually be distributed among many databases, but the publisher must be able to present metadata about the actual content to the DOI resolution system in order to communicate the location of the content at the publisher's site. The publishers will take the responsibility of maintaining the content directory over time such

as publisher restructures the files; moves the files to a new server; or even the change the ownership of the content.

When the DOI information is transferred to the publisher, it will look-up the DOIs from bibliographic data, citation linking, and find the contents. For subscribed users, the full-text article or downloadable document will appear on the screen, otherwise, information on how to obtain the content or other related data may appear.

DOIs and Metadata Standards

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage the characteristics of an information resource**.** It is a component of data in which describes the data or "data about data." The metadata describes who, what, when, where, why, and how about a data set. It provides the essential link between the information creator and the information user. Each DOI is associated with a series of metadata, a set of bibliographical and commercial information concerning the content (title, author, publication date, copyright, price, etc.). Answers for a number of basic questions about the identified resource (kernel metadata elements, a minimum level of publicly available structured metadata) should be known by the RA at the time the DOI name is issued and is becoming mandatory for all DOIs that are registered. Some of these data include an identifier associated with the entity from a legacy identification scheme, such as an ISBN (identifier), a name by which the entity is known (title), the primary type of intellectual property (type), the sensory mode such as visual/audio (mode) and primary agent and etc. Therefore when the publishers register DOIs, they also register kernel metadata that associated with the DOIs. In order to

register these kernel metadata, publishers must locate it within its internal system and they are responsible to maintain these DOI associated metadata over time. Digital Object Identifiers and their associated metadata are key ingredients in enabling the building of an infrastructure to support a cross-publisher database of journal articles and citations based on a distributed production model.[3] In order for the DOI to be able to fulfill its wider potential of providing the basis for a full range of services relating to intellectual property in the network environment, metadata becomes an essential component of the DOI System as a whole[4].

Use in the Library

Publishers today are investing in linking technology provided by the CrossRef (one of the DOI Registration Agents, http://www.crossref.org/) to benefit those who search the journal literature using the DOI as an open standard to make citation linking efficient and reliable. Publishers like ScienceDirect, Wiley InterScience, Blackwell Synergy, SpringerLink, and others had linked their documents via the CrossRef. The demonstration of these linking articles can be found at http://www.crossref.org/03libraries/18gallery.html. As the reference-linking network for scholarly and professional publishers, CrossRef linking enables the user to navigate the literature, moving from one article or idea to another across journals and publishers. CrossRef is a full-scale implementation of the DOI to date with coverage of mainly scholarly and professional research content, Journal articles, books, conference proceedings, etc. According to Amy Brand, CrossRef's director of business development, CrossRef is "a publisher membership association for collaborative technologies, and

official DOI registration agency, and a cross-publisher citation-linking network."[5] To date, over 2300 publishers and societies have participated in CrossRef's collaborative service, enhancing the value of millions articles from thousands journals. As of this writing, more than twenty-four million DOIs had been registered in the CrossRef system since its inception in early 2000 with more than 15,000 journals. The best news for the libraries is that CrossRef access is FREE for libraries! Yes, there is no end-user charge associated with their use. Other benefits for libraries include increasing usage of acquired resources; expanded access to content not owned; and enhanced localized linking. Even though the total number of participating libraries is 1,216 as May 3, 2007, there is still a long way to convince libraries to participate. In US alone, there were 9129 public libraries in the 50 states and the District of Columbia in fiscal year 2001.[6] According to National Center for Education Statistics (2005), there were 4140 colleges and universities in US in 2003, about 90 percent of them had their own academic library. That is about 3700 academic libraries in the US. Compare to the total libraries that could be part of this development, the participated library is still a very small portion.

On October 30, 2006, CrossRef announced the launch of its freely available Simple-Text Query service to facilitate DOI look-up for researchers and publishers. Libraries should take the advantage of the CrossRef and provide more efficient services for their patrons. For example, library can retrieve DOIs and metadata to make persistent links to full-text works online and use DOIs for online reserves by embedding in reserve URLs. Scott Warren has demonstrated in the construction of deeplinked e-reserves and provided examples of how such linking could take place (DOI: 10.1300/J124v22n04_01).[7]

CrossRef publishers support libraries not only by allowing patrons to move seamlessly

through the relevant body of research but also expanding access to content in which they do not own. Therefore, it will increase usage of acquired resources and enhancing end-user satisfaction with a new level of service. Libraries should join CrossRef and be part of these exciting developments. For more information on how to join CrossRef and the benefits of membership, please visit http://www.crossref.org/03libraries/28benefits.html. The benefits for libraries include increasing usage of acquired resources; expanded access to content not owned; enhanced localized linking and DOI and metadata retrieval privileges at no cost.

Benefits of DOIs and the Issues Related

The best feature of DOI is persistence. The DOI interlinking across that publisher's content will transport seamlessly, with the content itself, to your site, enabling users to experience the same richly and contextually interlinked content as if they were on the publisher's site.[8] Perhaps the most immediate and exciting benefit to a publisher of adopting the DOI, though, is that by making the DOI available to end-users, the publisher can effortlessly turn any pre-existing identifier into an persistent, actionable identifier with an efficient, scalable, Internet-based resolution and routing system behind it.[9] This makes easier for the publishers to manage their data/contents and keep tracking the items in their system. By using DOI, even if the content changes hands between publishers and/or vendors. Publishers and/or vendors only need to make one change in one place to make sure that all links point to the correct location.

DOI is an open system and anyone can build systems which integrate DOI into its local environments. For example, library can integrate DOI to access its locally licensed

materials. As mentioned before, the DOI system is fully managed via registration agencies and can be applied to any form of intellectual property at any level of granularity. The system interoperates with other data from other sources. For the end users, information can be accessed seamlessly cross the Internet.

Some issues related to the DOIs include intellectual property management. Who own the copyright of the material? How to deal with manifestations (the same content in another form, e.g. word format or PDF format) and version (e.g. edited or translated)? Should different DOIs be assigned for each manifestation and for each version? What metadata standard should be used when assigned DOIs? Even though there are some issues regarding DOI require further discussion, the DOI has been an extraordinary venture for the publishing industry and their customers and the end users have benefited tremendously from it.

Summary

The Digital Object Identifier (DOI) system not only provides a unique identification for digital content, but also provides a way to link users to the right material holders via an automated digital environment. The persistent, unique and reliable identifier along with the resolution system provides a perfect solution to manage the intellectual content in a digital environment. In the transition from paper to electronic publishing, the DOI has brought publishers and technologies together and overcome the significant obstacles of managing electronic content and libraries should take advantage of this opportunity to benefit their patrons and users. Some scholars state that, in the near future the value of a publication will be determined by the number of links to it. The widely adoption of the

DOI will not only benefit the scholarly community but also will enhance the e-commerce market for their digital asset management. Even though some issues are still around DOI, such as copyrights protection for DOI; metadata standards via assigning DOIs and missing links, we should continue to educate and market on DOI and to encourage the development of new applications and ensuring that the DOI fits into the emerging web infrastructure.

Notes

(1) See http://www.doi.org/

(2) See http://www.doi.org/sun_pap2.html

(3) "Discussion Paper: Role of the Digital Object Identifier in the Journal Portal Initiative and its ramifications/requirements in a distributed environment" (see http://www.hil.unb.ca/Texts/burk/sshrc_grant/Web/paper1.htm)

(4) David Sidman, "The Digital Object Identifier: The Keystone for Digital Rights Management" (see http://www.contentdirections.com/materials/SIIA-DOIandDRM-DavidSidman.htm)

(5) Marla Misek, "CrossRef: Citation Linking Gets a Backbone," *Econtent*, 27, no. 6 (June 2004): 44 – 45.

(6) Adrienne Chute, P. Elaine Kroe, "Public Libraries in the United States: Fiscal Year 2001," *Education Statistics Quarterly*, 5, no. 2, (2003): 147-149

(7) Scott A. Warren, "DOIs and Deeplinked E-Reserves: Innovative Links for the Future," *Technical Services Quarterly*, 22, no. 4 (2005): 1 – 17.

(8) Hal Espo, "Using the Digital Object Identifier (DOI) for Intranet Linking" (see http://www.intranetstoday.com/Articles/Default2.aspx?ArticleID=5501&IssueID=182)

(9) David Sidman, Tom Davidson, "A Practical Guide to Automating the Digital Supply Chain with the Digital Object Identifier (DOI)" (see http://www.contentdirections.com/materials/PRQ-CDIPracticalGuide.htm)

1
2
3
4
5
6
7
8
9