DIGITAL OBJECT IDENTIFIER (DOI): AN ISBN FOR THE 21ST CENTURY

Sunitha T
Sreejaya P
Indian Institute of Management
Kunnamangalam (P.O)
Kozhikode - 673 571
Kerala
Email: sunitha@iimk.ac.in
       sreejaya@iimk.ac.in

The Internet represents a totally new environment for the exchange of scholarly literature and it requires new enabling technologies to protect both customers and publishers. Systems will have to be developed to authenticate contents to ensure that what the customer is requesting is what is being delivered. At the same time, the creator of the information must be sure that copyright in the content is respected and protected. While considering the requirements of the new systems, publishers of books and journals internationally realized that a first step would be the development of a new identification system to be used for all digital content. The Digital Object Identifier (DOI) system not only provides a unique identification for contents, but also provides a way to link users of the material to the right holders to facilitate automated digital commerce in the digital environment. The DOI, an initiative taken by the Association of American Publishers, has grown into a prototype system, which supports URN to URL resolution and potential copyright management, document delivery and electronic commerce functions. In principle, the DOI system is fairly simple. It is based on the proposal that publishers should uniformly adopt formal or standard numbering schemes for their online products. The existing legacy systems such as ISBN, ISSN etc., should be used as the basis for this numbering. Such numbered products and their online addresses will be registered in an online directory or database which will automatically route queries about a product to the correct location of that product, where publisher's response screens will automatically offer the enquirer a variety of commercial transaction options such as viewing, purchasing, licensing or further routing. (For example: from article abstract to full text article content). DOI is a tool for persistence, multiple resolution, stability, security and authentication, distributed administration and internationalization, with a lightweight resolution method and an open protocol. This paper emphasizes on the origin, structure, application and advantages of DOI and its implications in digital rights management.

INTRODUCTION

The future of content industry is now clearly tied to the future of the Internet requiring a fundamental shift from physical to electronic dissemination. All types of intellectual property (e.g. books, music, journals, video, software) will be involved in this migration. One of the key issues in the move from physical to electronic distribution of contents is the evolution of a common technical and procedural infrastructure to identify or name pieces of content in a digital environment. The rise of electronic publishing poses new challenges to identification schemes and version control. However, the first and foremost challenge is the sheer volume and growth rate of electronic information. Established identification schemes such as ISSN and ISBN are not scalable due to their fixed number of positions. Another major challenge is the issue of granularity. Electronic publishing makes it easier to repackage information in a variety of ways and electronic distribution makes it easier to customize information packaging to individual end user needs. Publishers and the trading communities have come up with several solutions to cope with the granularity issues. The Internet Community has recognized the need to develop a Uniform Resource Identifier (URI) architecture for the operation of resource discovery systems on the Net that is more comprehensive and persistent than the URL addressing scheme alone. Working groups have been set-up to develop this architecture. The basic building blocks of this architecture are existing standard for location (URL), a standard for identification (Uniform Resource Name) and a standard resolution method for combining resource identification and location.

One promising development that builds on URI building blocks is the Digital Object Identifier (DOI) system. The DOI is an initiative under taken by the *Association of American Publishers*. It has grown into a prototype system which supports URN to URL resolution and potential copyright management, document delivery and electronic commerce functions. The versatility of the DOI system is indeed its most remarkable characteristics, combining a variety of creative online application with the potential for the standardization of document identification and retrieval which has hitherto been lacking in the digital publishing universe.

The DOI system also promises profound benefits for libraries and other information organizations in their quest to move from printed to digital or virtual collections, such as facilitation of improved WEBPAC, management of online resources and of navigable reference citation possibilities.

WHAT IS DOI?

The Digital Object Identifier (DOI) [1] is a system for interoperably identifying and exchanging intellectual property in the digital environment. It provides an extensible framework for managing intellectual content in any form at any level of granularity for linking customers with content suppliers, facilitating electronic commerce, and enabling automated copyright management for all types of media. The DOI is a unique, persistent, managed, international public identifier, which is assigned when the article is accepted for publication.

According to Norman Paskin [2], DOI is a unique identifier of a piece of content and a system to access that content digitally - in essence, an ISBN for the 21st Century.

ORIGIN OF DOI

One of the key issues in the move from physical to electronic distribution of contents is the evolution of a common technical and procedural infrastructure to identify, or name pieces of contents in a digital environment. The DOI is thus an enabling technology for providing services and for protecting contents in the digital environment. It was developed to ensure precise identification of content and thereby enable precise management of rights. The international content producers realized that the first step was the development of an identification system that was persistent and could link content to multiple locations and services. Some initial experiences with the formal numbering of commercial online documents was gained with the creation of the Publisher Item Identifier (PII) which was drawn up in 1995 by a group of Scientific Publishers consisting of the *American Chemical Society, American Institute of Physics,* Elsevier Science and the *Institute of Electrical and Electronic Engineers.* These publishers compiled a list of characteristics for an online document identifier which stipulated the exchange in the electronic environment, compatibility with existing standards, future extension and uniqueness.

Following this, the *American Association of Publishers* [3] agreed in October 1996 on primary need for unique product identifiers in order to exercise control over their products in cyberspace. A number of international organizations like *International Publishers Association* (IPA), *Book Industry Communication* (BIC), *Internet Exploring Task Force* (IETF), *National Information Standards Organization* (NISO), ISBN International and *Authors Collecting and Licensing Society* (ACLS) also began to monitor and support the issues of standardized document identifiers.

Work on the DOI began in 1996 following a request for proposals for technologies to identify content issued by the *Association of American Publishers Inc* (AAP) Enabling Technologies Committee. The Committee selected a joint submission from the R.R. Bowker Company and the *Corporation for National Research Initiative* (CNRI) and the DOI system was officially launched in the second half of 1997. Since 1998, the DOI has been managed by the International DOI Foundation.

DOI Structure

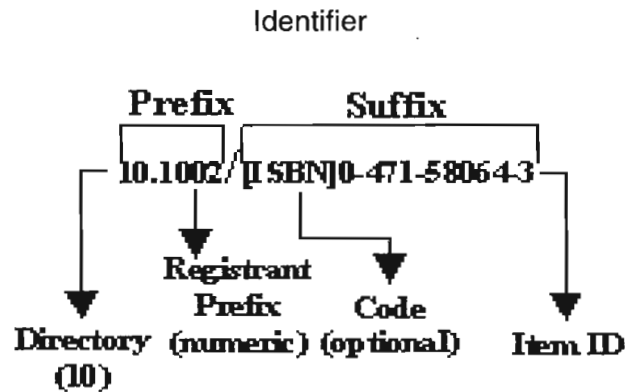The DOI system has three parts, the Identifier, the Directory and the Database.

Identifier

Fig 1

Identifier Components

10.1002/0-471-58054-3

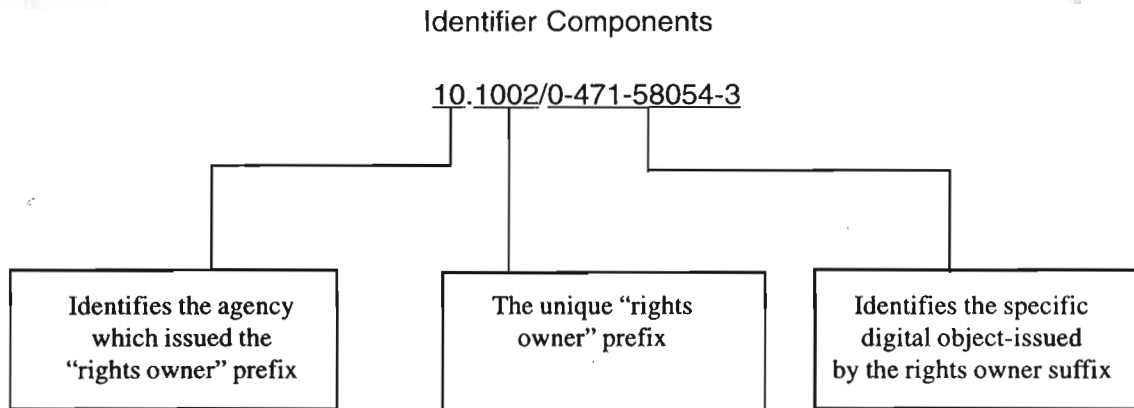| Identifies the agency which issued the "rights owner" prefix | The unique "rights owner" prefix | Identifies the specific digital object-issued by the rights owner suffix |

Fig 2

Identifier

The identifier is made up of two elements. The first element is prefix and second is suffix (Figures 1 & 2). Prefix comprises two components. Directory code and a registrant's code. Prefix (1002 in this example) is a number that corresponds to the publisher (the rights owner) which is assigned by the directory manager. All prefixes begin with 10(Directory code), that can denote the directory that assigned the prefix to that publisher followed by a period (.).

The second element, suffix is followed by a slash mark (/). This is the designation assigned by the content producer to the specific entity being identified. Suffix begins with an optional code designator in square brackets which can be helpful as a signal to software. Many publishers have elected to use recognized international standards for their suffixes. In the Figure1 ISBN for the International Standard Book Number follows the actual item identifier (0-471-58054-3).

The suffix can be assigned to entities of any size of granularity – book, article, abstract, chart, album, song- or any file type-text audio, video, image or software. The suffix can be as simple as a sequential number of a publisher's own internal numbering system.

DOI system involves a central directory by which it works. This directory resolves them to internet addresses which are internet URLs. When a user clicks on a DOI, message is sent back to central directory where the current web address associated with that DOI appears. This location is sent back to the user's internet browser with a special message telling the system to "go to the particular internet address". The user simply sees the content itself, or further information about the entity as well as the information on how to obtain it.

When the object is moved to a new server or the copyright holder moves the product line to another company, one change is recorded in the directory and all subsequent users will be sent to the new location. The DOI directory is a distributed database based on CNRI's handle system that provides a mapping from each DOI to a URL. The directory as a whole is maintained by the directory manager(s) though individual entries are inserted and updated, by publishers once they have obtained a DOI prefix facilities for bulk loading DOIs in to the directory.

Database

The publisher/rights-holder maintains a database that contains the actual content, plus information about the content. It is anticipated that for many publishers, the URL held in the DOI directory will point to a 'response screen' rather than to the digital object directly. Such a response screen will be maintained by the owner or publisher. It might include the actual content or information on where and how to obtain the content or other related data.

TECHNOLOGY

The underlying technology for the DOI system is the Handle system developed by the *Corporation for National Research Initiatives* (CNRI). This system is a distributed scalable system based on open protocols, which manages digital objects as first class entities.

COMPONENTS OF DOI

DOI System is made up of four primary components, namely, Enumeration, Description, Resolution and Policies.

Enumeration

Each DOI is a unique "number", assigned to identify only one entity. It is important for the integrity of the system that the same number is not used twice to identify different things. The DOI is designed in such a way as to make it as simple as possible for anyone to name uniquely any item of intellectual property — tangible or intangible, in physical or digital form. In use, the DOI is an alphanumeric string that contains both characters and numbers

Description

Metadata is an essential component of the DOI System. Rich descriptive metadata, coupled with an intelligent search system, enables accurate and efficient retrieval of online documents. To set appropriate limits on the amount of metadata, a very small "kernel" of metadata will be used with every DOI as below:

- Identifier: an identifier associated with the entity from a legacy identification scheme (where such an identifier exists).

- Title: a name by which the entity is known

- Type: the type of intellectual property entity that is being identified (an abstract "work", a digital or physical "manifestation", a performance)

- Mode: the sensory mode through which the intellectual property entity is intended to be perceived (visual, audio, audiovisual)

- Primary agent: the identity of the "primary agent", normally the first-named creator of the entity.

- Agent role: the role that the primary agent played in the creation of the entity.

This metadata will be available to any user of the DOI System, to enable them to find a basic description of the entity that any particular DOI identifies. This basic description will allow the user to understand some basic things about the entity.

Resolution

The DOI System is different from most other identification systems in being actionable. A DOI

on the Internet can be "resolved", leading the user of a DOI to any piece of data that is Internet-accessible. A resolution system takes a URN and returns a list of services or instances of the information identified by the URN, commonly one or more URLS. Resolution here is used to mean the act of submitting an identifier to a network service and receiving in return one or more pieces of current information related to the identifier. In the case of the Domain Name System (DNS), the resolution is from domain name to a single IP address, which is then used to communicate with that internet host.

The DOI, though, can be used to identify classes of intellectual property — abstract "works", physical "manifestations", performances - cannot be directly accessed in a digital file. Even when the DOI does identify a digital file, this will not always be the most appropriate or useful data for the DOI to resolve to. Even if there is no current location for a digital file, it might still be useful to know what it represented, or who owned it, or search for it elsewhere. Even if we have a location, we might want to offer other resolution results. Therefore it is very important to distinguish what the DOI identifies from what the DOI resolves to. They may be the same thing, but they will often be very different.

Policies

Like any other system DOI also requires rules for its operation. The DOI can be distinguished from other identification schemes, in particular from other implementations of the Handle System, by its policies. These policies ensure that the DOI System provides reliable and predictable results to the user. Reliability and predictability can be regarded as the metrics of success of the quality assurance regime provided by the DOI's policy framework.

ADVANTAGES OF DOI

The DOI has following advantages:
1. It identifies the entity, not the location, whenever the location changes, the DOI remains the same, persistency can be assured every time.

2. The usage of DOI is free and it can be used by anyone.

3. It is a fully managed system.

4. DOI can be applied to any form of intellectual property, at any level of granularity [4]. DOI can work with the existing identifier system such as ISBN, ISSN, ISRC etc and make them actionable. The ISBN International is incorporating the administration of DOI function into the existing worldwide ISBN network of product numbering agencies to promote DOI as a formal information standard.

5. DOI is not limited to the current environment, it is applicable in any digital environment, the concepts are designed to be useful in any future Internet protocols.

6. DOI system provides a model for online copyright developments which will impact on the way online intellectual property is protected in the future.

7. Metadata is managed in a controlled way. The DOI numbering string could be adapted for metadata uses to become the most reliable reference and retrieval link between the source document and abstracted or indexed descriptions of the source contents.

8. DOI can point to the online commercial transactions. Clicking on the DOI buttons for article references offers menus from which users can choose to view the journal table of contents, the full articles, or order copies of the article.

DOI AND E-BOOKS

The DOI is like a supercharged bar code for Internet-based resources such as digital content published online. It uniquely identifies digital objects and provides permanent links to the publisher and/or to any related services the publisher wants to enable the automation of the supply chain, thus facilitating of online

transactions of all kinds including e-commerce, rights management, and digital distribution.

The endorsement of DOI by book publishers is a big boost to DOI, which until now has been embraced by the journal community and was largely ignored by everyone else. DOI-EB is a project sponsored by the International DOI Foundation and it has participation of several large book publishers like McGraw-Hill, Random House/Bertelsmann, John Wiley & Sons, etc., and a number of technology companies including Adobe, Microsoft, iCopyright, NetPaper, etc [5]. Its purpose is to demonstrate the commercial viability of the DOI system for the Book industry, in the same way that a previous project, "DOI-X", demonstrated the key real-world application that led to the DOI's adoption by the scientific journal industry. The e-book project addresses that the publishers can assign DOI numbers to parts of books so that they can sell individual chapters to be downloaded or combined with other material into college course packs or print -on-demand books.

Registering DOIs will allow publishers to sell e-books, individual chapters and other innovative forms of content. It will also allow them to continue selling traditional and physical books. In addition, the DOI multilink facilitates the availability of free excerpts, exposure to book reviews, access to the publisher's catalog page for additional related information, and sales across a publisher's distribution chain regardless of format, all directly from within Adobe Acrobat eBook Reader or Microsoft Reader on the reader's PC.

DOI AND DIGITAL RIGHTS MANAGEMENT

Digital Rights Management (DRM) is a new acronym on the intellectual property scene, but its core concept has actually been around for a number of years [6]. DRM systems are a response to the heightened expectations of information professionals, among others for more efficient transactional licensing of digital contents. Digital Rights Management systems are best understood as databases that

streamline the complex relationships and transactions among rights, works and parties.

When publishers created the Digital Object Identifier in 1996, they had two goals in mind which are:

(1) facilitating the creation of an e-commerce market for digital contents, and

(2) facilitating the creation of solutions for digital rights management [7].

E-commerce of intellectual property involves transacting intellectual property not entities, or the right to the intellectual property entities. The commonly used phrase "Digital Rights Management" in fact implies both, management of Digital rights and management of rights.

Digital Rights Management (DRM) goes well beyond mere copyright protection in the negative sense of preventing piracy. It has many affirmative benefits as well, which in theory make it a tremendously powerful marketing vehicle. It can allow limited access but then requires purchase if the customer wants to print, copy /paste or forward to others. DRM can even involve payment but the users have to register their name and email addresses as a precondition of obtaining access.

But all of these capabilities still require successful interoperation of different systems from the publisher's internal production systems, to the publisher's web publishing systems to the DRM wrapper which encrypts or seals the content to the hosting provider which stores the content for distribution to the online bookstore. This in turn stores the content for sale to the rights clearing house, which checks the customer's authority to access it and further to the e-commerce vendor who takes the customers credit and payment and to the issuer of the key or permit which unlocks the content.

The DOI offers DRM a tremendously powerful tool in terms of actually preventing piracy. Many DRM solutions today fail at the point when the

content is legitimately opened by a legitimate customer at which point the content may be easily copied. Many DRM solutions also fail because it is almost impossible to prevent hacking of encryption solutions as long as content itself, the locking mechanism and the key are all located somewhere within the content or on the same reading device.

The DOI assists these solutions in a number of ways. One of these is that the DOI represents an "actionable identifier", not just a unique number in the abstract, i.e., it provides a permanent, reliable link back to the content owner's server. This ability of the DOI to "phone home" for access (e.g to obtain the server half of a required key pair) is a powerful tool for DRM.

The very permanence of a DOI based link represents another way that the DOI helps DRM. The DOI can be embedded in the content's secure wrapper at the outset of publication, because even if the content is then passed along from user to user over a number of years, the embedded DOI will still take the newest customer to a valid place on the internet in order to successfully complete a DRM transaction and get access. Again that place may be the content owner's website, or another website that the content owner may designate as its preferred right clearing house or e-commerce vendor. But whatever target URL is, the DOI will always point somewhere successfully even if that target URL has to be changed over time.

With widespread adoption of DOI by the publishers in terms of assigning DOIs to their content and in parallel, by DRM vendors, content management /web publishing system vendors, online bookstores, and all other distribution partners should support the DOI as their standard object identifier for digital content and this will lead towards the establishment of a vibrant e-commerce market place for digital content.

DOI AND LIBRARY

The DOI initiative has prompted significant amount of discussions in the library community.

With the increase in the use of electronic information services in the library particularly with the dramatic increase in the use of the electronic journals, there is an urgent need in libraries for solutions that link the information resources in a meaningful way for the end user and that optimize the use of these resources. Such linking solutions are now available for libraries supported and assisted by DOI . The linking solutions built around DOI include SFX and Cross Ref [8]. The DOI, CrossRef , SFX and OpenURL are complementary frameworks and components that can be integrated. Collaboration between the SFX community and the DOI community is underway to integrate the SFX framework and the DOI framework in order to enable localized resolution of DOIs. The CrossRef linking solution is gaining widespread acceptance and now has more than 200 member publishers, all of whom have agreed to enable linking of a publisher's reference citations to the online content that those references cite, typically located on a different server and published by a different publisher. This linking is based on the DOI. In this CrossRef linking framework, the publisher controls resolution of CrossRef links and thus the framework does not readily allow for context-sensitive linking. It does not address the "appropriate copy" problem and this is of great concern to many libraries.

SFX launched by ExLibris is an OpenURL compliant open linking solution for libraries. It is an institutional service component that provides libraries with an independent means of offering seamless interconnectivity between their ever-increasing collections of information resources. With SFX, libraries can define a range of extended services (different kinds of links). SFX enables the localisation of services and uses the appropriate links; and allows for standardisation of services across resources as defined by the librarian. The SFX link server in the library can be configured to reflect its collection.

Clicking on the SFX button will provide a number of service options for the user. Users clicking on a Crossref (DOI) link embedded in an article

reference will be directed, via the DOI resolver, to the full text of the article at the publisher's site. If a library does not have subscription to the journal, the user may be denied acess to the full text- or given an option to purchase the article. If the library has an SFX link server, then the DOI server will recognize this and redirect the DOI- as an OpenURL- to the library's SFX server. The SFX server will use the CrossRef database to retrieve the article metadata and can use this to determine the appropriate SFX service.

Apart from these, if universally implemented, the DOI system could provide many more benefits for libraries and other information organizations.

Information organizations wishing to expand their print collection by "cyberstacking", or integrating virtual or electronic resources into their Webpacs, will be able to insert the DOI number into their 856 MARC electronic resource catalogue description records as the most reliable retrieval element for the resource owned or licensed by the library [9]. The DOI number may then be hot linked in their Webpacs to provide persistent access to that resource. Since the location of the resource is automatically updated by the DOI directory itself, this will minimize the daily link checking which is currently standard practice for catalogue and WebPac maintenance. The use of the DOI in metadata packages should lead to more efficient and standardized methods of document reference and retrieval in the information intermediary industry, and may prove particularly beneficial for the use of indexing and abstracting services supporting interlibrary lending activity and document delivery.

CONCLUSION

DOI system will help to satisfy many of the needs of the major publishers both in supporting commercial interactions and in protecting intellectual property. It provides a creative opportunity for publishers to introduce systems that protect them from serious financial losses

accruing from the misuse of digital material. It is obvious that DOI would be best for publishers to collaborate and corporate with their clients such as libraries to establish DOI system policies and applications that adequately support both distribution and access.

Despite the benefits of DOI, there are issues concerning DOI system. Presently, only established commercial and society publishers are allowed to purchase publisher prefixes and issue DOIs. The cost of overheads in assigning DOIs and keeping up with many detailed housekeeping problems associated with their day-to-day maintenance might turn out to be expensive. Many other issues remain unresolved or have no easy solutions. There are many questions arising about the version, format, stage of document etc. It appears for the present that besides its limitations, DOI remains the most ambitious identifier in the history of the world.

The aim of DOI is to put in place an enabling technology, which will enable publishers and users to overcome the significant obstacles, in the way of managing electronic content while preserving rights. The DOI offers the potential for improved access by using persistent, managed, identifiers and mechanisms for their use in services offered by publishers to libraries and readers.

REFERENCES

1. http://www.doi.org

2. LISA (Stickley). Norman Paskin discusses the Digital Object Identifier. *OCLC Newsletter.* 239; 1999. Available at http://www.oclc.org/oclc/new/n239/research/01research.htm

3. ROSENBLATT (Bill). Solving the dilemma of copyright protection online. *Journal of Electonic Publishing.* 3, 2; 1997; 285.

4. BIDE (Mark). In Search of the Unicorn: The Digital Object Identifier from a User Perspective".*BNBRF Report 89,* Book Industry Communication, February 1998. Available at http://www.bic.org.uk/unicorn2.pdf.

5. WALTER (Mark). E-book Project Highlights Role of DOI in selling Digital Content. The Seybold Report.

1,6;2001. Available at www.seyboldreports.com.

6. DAVIS (David D). What digital right management means today. *Computers in Libraries.* 21, 6; 2001; 36-39.

7. DAVIDSON (Lloyd A) and DOUGLAS (Kimberley) DOI: Promise and problems of scholarly publishing. *Journal of Electronic Publishing.* 4,2; 1998; 293.

8. WALKER (Fenny). CrossRef and SFX: Complimentary linking services for libraries. *New Library World.* 103, 1174; 2002; 83- 89.

9. MAREE (Jean). The digital object identifier system and its applications in the online information industry: An overview for Southern African Information Professionals.(1998) Available at http://www.unisa.ac.za/library2/afdeling/tdienste/doi.html