Electronic Scholarly Journals: A Review of Technical Issues in Digital Environment

Golnessa Galyani Moghaddam

Faculty
Dept. of Studies in Library & Information Science, Shahed University
Tehran, Iran
E-mail: g_galyani@yahoo.com

Shalini R. Urs

Professor
Dept. of Studies in Library and Information Science, University of Mysore
Mysore, Karnataka, India
E-mail: shalini@vidyanidhi.org.in

Abstract

Scholarly journals are known as the most important medium for scholarly communication since long time back. As technology transforms the flow of information and idea everywhere, it changes the nature of scholarly communications and publishing of scholarly journals as well. The electronic scholarly publishing rapidly produced an expectation, among researchers of the availability of articles at their desktop, rather than the previous scenario of visiting the library to read a print journal issue. There are lots of technological improvements in electronic journals publishing. The present paper looks at some of technical issues in electronic publishing such as DOI, DOI-X, CrossRef, Citation/Reference Linking, OpenURL, SFX and MetaLib which are being used in the World Wide Web.

Keywords: *Electronic journals; DOI; CrossRef; Citation/Reference linking; OpenURL; SFX; MetaLib*

Introduction

Electronic scholarly journals are media for scholarly communication in digital environment with the ability to provide information much timelier than print journals. Although, the reader is dependent on access to a computer, rapid distribution directly to the desktops of subscribers ensures timely delivery in geographically remote locations. Presence of electronic journals needs infrastructures technological improvement in networking and computer and software. There are lots of technological improvements in electronic publishing. The purpose of this paper is to give a brief explanation of the following technical issues: DOI, DOI-X, CrossRef, Citation/Reference Linking, OpenURL, SFX and MetaLib.

2006/10/06 received; 2006/11/26 revised; 2006/12/11 accepted

Background

In the print environment, the capabilities and functionality of scholarly journals were limited while information technology brought many possibilities to them which were impossible before. With the help of technological improvements, electronic scholarly journals are having many new features and functionality in the online environment. Morris (2006) pointed out five new capabilities for electronic journals. According to her, one key feature of online publications is the ability to search for articles — at a minimum within one journal, and frequently across multiple journals.

Another feature commonly added to online journals is that of linking. This refers to the ability to go from one place online to another, at a click of a button — the most common example being the link from a citation within the body of an article to the relevant reference at the end of the article, and from there to the actual article which is being referenced. Reference linking in scholarly journals is now increasingly being undertaken using DOIs, which are unique identification codes assigned to online articles, and which provide permanent links for the user. This is one of the international standards being developed to assist different systems to communicate and link with each other, to help with visibility and discovery of online material. The methodology of using DOIs in scholarly journals is being managed by an organisation called CrossRef (see the DOI in this paper).

Electronic publication also makes it possible to include material which researchers would not include in a print journal, either for space reasons or because the material is difficult, costly or even impossible to print. Research articles online can be accompanied by the full data-sets from which the results were reached — ideally in a form which readers can manipulate for themselves. Colour illustrations may be prohibitively expensive in a printed journal, but present no such problems in an electronic journal.

Moving images, sound, or animations may also be included, although relatively few authors automatically think of providing such material at present and the technical complexities may outweigh the benefits.

The last additional capabilities which online publishing can offer may be of particular relevance to author, as the publisher, and not so much use to the user. These may involve the ease with which authors can upload their content, and streamline their working practices. One particularly important new facility which online publishing can provide, is a tool to measure and monitor online use. This can also be of great value to librarians so they can see how valuable their journal is to their users. Most commercial hosts will automatically offer sophisticated

usage statistics (Morris, 2006).

It may be noted that linking is generally rated very highly by readers, who find it an invaluable way of enhancing their research; however, other additional features are much less highly valued (Baldwin and Pullinger, 2000).

Technical Issues of Electronic Journals

Now, we look at some technical issues in digital environment with refer to electronic journals to know more about their new features and functionality.

DOI (Digital Object Identifier)

The Digital Object Identifier (DOI) is a system for persistent identification and interoperable exchange of intellectual property on digital networks. It provides an extensible framework for managing intellectual content in any form, at any level of granularity and in any digital environment. The International DOI Foundation, a non-profit organization, manages development, policy, and licensing of the DOI system to registration agencies (http://www.doi.org).

The DOI is a unique identifier for articles (and, indeed, other online content), which is associated with the article's current location on the web; when the location changes, the DOI stays the same, but the database showing the current URL with which it corresponds is updated once — this is far more reliable than hoping that all instances of the URL can be updated (Baron, 1997).

There are formal international standards for the operation and implementation of the DOI. The publisher needs to obtain a unique number (prefix), and then needs to assign a unique number (suffix) to each article or other item. The combination (prefix + suffix) must then be registered with a DOI registration agency (see Figure 1). There are costs associated with this, but they vary with each registration agency. Within the scholarly publishing industry, the agency used by most publishers is CrossRef. CrossRef has developed an application of the DOI system to simplify linking from one journal to another (particularly from a reference at the end of one article to the cited article in another journal). This is increasingly used by Western publishers.

By integrating an identifier into a DOI, the identifier becomes actionable as a standard hyperlink and can function in DOI applications across different platforms. A variety of different identifier systems become readily interoperable when incorporated into DOIs. DOIs may be assigned to ISBN entities (books) to achieve this; DOIs may be used to identify related entities or linked material in any form.

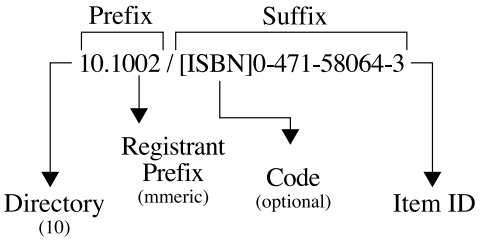

Figure 1 The DOI number composed of three parts

Source: The International DOI Foundation

The DOI is functionally similar to a Uniform Resource Locator (URL) in that a user can click on it and go directly to the DOI Directory, which in turn seamlessly reroutes the user to the source of information corresponding to that DOI. Unlike a URL, the DOI can easily be rerouted. A rights holder who purchases rights to a work from another rights holder can update the Directory information to ensure that future clicks are routed to its system.

To give an example of DOI, see the following information for a paper which is being published by Haworth Press in *The Serials Librarian*:

Journal title: *The Serials Librarian* (ISSN: 0361-526X)
Volume: 51 Issue: 3/4
Publication Date: 2006
Title of paper: Scholarly Electronic Journal Publishing: A Study Comparing Commercial and Nonprofit/University Publishers
Page Range: 177 - 195
DOI: 10.1300/J123v51n03_12
Author: Golnessa Galyani Moghaddam

A look on this issue of *The Serials Librarian* journal reveals that the first part of DOIs for all papers are same. The only difference is in the last number. For the above paper, the last number is 12 and refers to twelfth paper which is publishing in this issue of the journal.

DOI-X

DOI-X is a prototype metadata database designed to support DOI lookups. The prototype is intended to address the integration of metadata registration and maintenance with basic DOI registration and maintenance, enabling publishers to use a single mechanism and a single quality-assurance process to register both DOIs and their associated metadata. It also contains the lookup mechanisms necessary to access the journal article metadata, both on a single-item lookup basis and on a batch basis, such as would facilitate reference linking. The proto-

type database was introduced and demonstrated to attendees at the STM International Meeting and the Frankfurt Book Fair in October 1999.

The DOI-X data format was specified in an XML Document Type Definition (DTD) and accompanying "rules document". The rules document provided DTD documentation and specific constraints that could not be expressed in XML (e.g., ISO date format; limitation of journal titles to 256 characters; etc.). The DTD was designed to capture in discrete records the metadata about the full-text of a journal article, an abstract, or a bibliographic record. Allowing deposit of DOIs and metadata for secondary database records could have enabled the creation of links to bibliographic records, possibly with abstracts, for older articles not yet put online by publishers.

CrossRef

CrossRef (http://www.crossref.org) is a publisher collaborative that operated by the Publishers International Linking Association, Inc. (PILA). CrossRef, which was established in 2000 by scholarly publishers as an independent, nonprofit entity, uses open standards. It's an official registration agency of the International DOI Foundation (IDF) and is the first full-scale implementation of the Digital Object Identifier.

CrossRef's general purpose is to promote the development and cooperative use of new technologies to speed and facilitate scholarly research. The specific CrossRef mission is to be the citation linking backbone for all scholarly information in electronic form. CrossRef is a collaborative reference linking service that functions as a sort of digital switchboard. It holds no full text content, but rather effects linkages through Digital Object Identifiers (DOI), which are tagged to article metadata supplied by the participating publishers. The end result is an efficient linking system through which a researcher can click on a reference citation in a journal and access the cited article.

CrossRef is a process, not a product. Each member publisher creates a DOI incorporating its own DOI prefix for each journal article, tagging it to article metadata and a URL. Records are assembled into batch file submissions to the CrossRef metadata database (MDDB) in a strict XML-based DTD format. CrossRef then registers each article DOI and URL in a central DOI directory. In a separate process, the publisher also submits the reference citations contained in each article to the reference resolver, a front-end component of the MDDB that allows for the retrieval of DOIs. The publisher can insert CrossRef links into any of an article's citations that point to content already registered in the CrossRef system.

Citation/Reference linking

Caplan (2001) defines citation or reference linking as "the ability to go directly from a citation to the work cited, or to additional information about the cited work," whether the source and accompanying destination are journal articles, Web sites, conference proceedings, entries in A&I databases, or even a link sent via e-mail from one colleague to another. In the scholarly electronic publishing community, citation or references linking tries to solve the obstacles and problems associated with linking among and between journal articles and bibliographic entries.

Grogg and Tenopir (2000) explained three kind of link. They addressed that links can be internal, contained within one service, or external, connecting documents or records provided by two or more services (Grogg and Tenopir, 2000). Internal linking occurs in aggregator services such as EBSCOhost's Academic Search Elite or Academic Search Premier, OCLC FirstSearch, Gale Group's InfoTrac, or ProQuest's Research Library. Primary publishers also employ internal linking in their own direct subscription services. For instance, reference links are available among journals published by Elsevier's ScienceDirect.

A prime example of the external linking model is linkages among secondary and primary publishers or links among abstracting and indexing services, primary publishers, and aggregators. Articles are housed on a different server than the bibliographic records. An A&I service functions as a navigational tool that then points users to the full text, via services such as SilverPlatter or Cambridge Scientific Abstracts (CSA) (Grogg, 2002).

A link is the specification of a relationship between a data source and destination where both the source and destination may expand to one of several places in a particular set of documents. These flexible relationships are usually coded explicitly by the Distributed Link Service (DLS) as a generic link, but they may be the by-product of a more complicated processing arrangement. For the purpose of citation linking this involves matching citations and bibliography lists in a document against a database of bibliographic information.

As well as recognising the citation, the agent has the problem of retrieving the cited paper, i.e. where to link the citation to. Extracting the citation is a more or less difficult task for the agent depending on the format in which the paper is held. Highly structured document formats based on SGML are in principle the simplest as all the separate components of the bibliography data are marked explicitly. At the other end of the scale, simple ASCII provides problems in recognising both the existence of a citation and the boundaries of its component

parts (for example, where the author names end and the title starts). HTML can provide extra clues implicit in the formatting markup (for example, titles may be rendered in italics).

OpenURL

The OpenURL provides a mechanism for encoding a citation for an information resource, typically a bibliographic resource, as a URL.

The citation is provided by either using a global identifier for the resource, for example a Digital Object Identifier (DOI), or by encoding metadata about the resource, for example title, author, journal title, etc., or by some combination of both approaches. It is also possible to encode a local identifier for the resource within the OpenURL. In combination with information about where the OpenURL was created, this allows software that receives the OpenURL to request further metadata about the information resource.

By clicking an OpenURL for a work, the user requests that the service component deliver extended services for that work. The service component takes the OpenURL as input and collects metadata and identifiers for the work. It can do this by directly parsing such information from the OpenURL and/or by fetching it using the metadata pointer that was provided in the OpenURL. This pointer can lead into the original resource or into another one. Once identifiers and metadata are collected, the service component will evaluate them and provide extended service links to the user. When the service component is appropriately tailored, these links will be sensitive to the context of the user.

An OpenURL comprises two parts, a BASEURL and a QUERY. The BASEURL identifies the OpenURL resolver that will provide context sensitive services for the OpenURL. The BASEURL is specific to the particular user that is being sent the OpenURL — it identifies the user's preferred OpenURL resolver. In many cases this will be the resolver offered by the institution to which the user belongs. Here is an example OpenURL:

> http://resolver.ukoln.ac.uk/openresolver/?sid=ukoln:ariadne&genre
> =article
> &atitle=Information%20gateways:%20collaboration%20on%20content
> &title=Online%20Information%20Review&issn=1468-4527&volume=24
> &spage=40&epage=45&artnum=1&aulast=Heery&aufirst=Rachel

In this example the BASEURL is <http://resolver.ukoln.ac.uk/openresolver/>, the URL of the UKOLN OpenResolver demonstrator service. The rest of the OpenURL is the QUERY, which is made up of a single DESCRIPTION of an article entitled "Information gateways: collaboration on content" by Rachel

Heery. The article was published in *Online Information Review* volume 24.

The original version of OpenURL, now referred to as OpenURL version 0.1, provided both a common linking syntax and a solution to the appropriate copy problem. The OpenURL concept was developed as part of a research project (called SFX, "special effects") by Herbert Van de Sompel and Patrick Hochstenbach at Ghent University in Belgium during1998-2000 (Van de Sompel, and Beit-Arie, 2001).

The OpenURL Framework for Context-Sensitive Services was endorsed as a NISO standard, Z39.88-2004 in April 2005 (NISO, 2005). This new standard has broadened the potential scope of OpenURL implementation beyond the scholarly information community, with the possibility of extension, by registration of new formats and profiles for new domains, as well as the introduction of an XML format, with its additional potential. Z39.88- 2004 includes details of the context of the link in a standard way, such as the user and the source of the link, opening up new possibilities of services appropriate to the user. The OpenURL Framework has separated the payload of an actionable link, which is the details of the reference and its context, known as the ContextObject, from the means of transporting it across the network, which is the OpenURL. This separation enables use of the ContextObject within other applications (Hodgson, 2005).

SFX

SFX is a context-sensitive link server from Ex Libris[1] that allows context-sensitive linking between Web resources in the scholarly information environment. SFX is OpenURL-compliant, in that it accepts an OpenURL as input from an Information Resource known as an SFX source (htpp://www.sfxit.com/).

An SFX source is a Web-based resource in which the user searches, and from which a user may link out to additional resources and services by clicking on an SFX button. The SFX button activates an OpenURL that sends metadata to the SFX server. A resource can only be an SFX Source if it provides an OpenURL.

There is relationship between SFX, CrossRef and DOI. As mentioned before, CrossRef is a consortium of primary publishers. CrossRef members use the Digital Object Identifier to link between their resources e.g. from the reference in an article from one CrossRef member to the full text of the article itself from

[1] The Ex Libris group is a worldwide supplier of software solutions and related services for libraries and information centers. The Company's flagship product, ALEPH 500 (Automated Library Expandable Program) is a market leader in the field of library automation for higher education as well as for public, national, and research libraries, consortia and national networks, and large corporations. Ex Libris has recently announced two new products: MetaLib, a front-end portal to scholarly resources, and SFX, a reference linking system supporting hybrid library environment.

another CrossRef member. The DOI, whilst offering a namespace-based linking solution, does not take into account the user's affiliation and therefore does not provide for context-sensitive linking services. However, when used within the SFX framework it can deliver such context-sensitive linking services. OpenURL/ SFX and CrossRef/DOI are compatible and complementary.

The SFX solution offers libraries flexibility and choice. Reference librarians can choose appropriate content from a range of information vendors and interconnect this content as desired. They can then provide links to services that they feel are appropriate for their end users. Reference librarians need not depend solely on the linking services defined by the information providers, on a specific set of identifiers (such as ISSN, SICI, or DOI), or on particular communication protocols (such as Z39.50 or http).

MetaLib

MetaLib is a standardized user interface and portal for users of today's and tomorrow's hybrid information systems. MetaLib is powered by two new Ex Libris technologies—the Universal Gateway and SFX. The Universal Gateway ensures accurate and target-sensitive searching and employs an intelligent analyzer to convert user requests into target specifications, and target data into user formats. SFX provides a host of contextual links to related information after performing an intensive, automated analysis of a document, including both the institutional and user environments.

MetaLib serves as a gateway to local and remote databases. MetaLib provides one gateway to both local and remote resources while allowing the user to remain within one familiar and uniform system. It enables an institution to provide its patrons with a unified interface to the diverse resources that it offers. MetaLib provides a hook that makes use of the institution's authentication mechanism. Once a user is authenticated and identified as belonging to a certain group of users, MetaLib's authorization mechanism, as defined by the institution, controls both the resources and the functions available to that user. MetaLib also provides for guest users, who may be members of the institution or may not be affiliated with it at all.

Discussion and Conclusion

Information technology had a profound impact on scholarly journals pub-lishing and brought many possibilities in digital environment which were impossible before. This paper looked at some technological issues in electronic pub-lishing such as DOI, DOI-X, CrossRef, Citation /Reference Linking, OpenURL, SFX and MetaLib.

Creation of DOI, a unique identifier for an article on the Web, seems to be a great advantage in electronic publishing and it lies at the core of other technological developments in electronic journals publishing. DOI are being used by many publishers now and it is more known than DOI-X. Prior to creation of DOI, URL as a unique identifier was arrived with the presence of the Internet, however, publishers enjoyed more from DOI in electronic publishing. DOI also led to other developments in digital environment. For example, CrossRef members, use DOI and URL both together to link between their resources e.g. from the reference in an article from one CrossRef member to the full text of the article itself from another CrossRef member. The OpenURL Framework developed from the SFX research work and is the architecture for localized and context-sensitive resolution of metadata and identifiers of referenced scholarly work into services. OpenURL can use DOI or URL separately and use them together. SFX, as a context-sensitive, gives librarians flexibility to find appropriate content from a range of information vendors and interconnect this content as desired. Generally, DOI is used in the literature along with CrossRef (such DOI/CrossRef) and SFX is used with OpenURL (such SFX/OpenURL). The DOI, CrossRef, SFX and OpenURL are not competitors but complementary services which can work together.

In spite of the above advances in electronic publishing, the current technical environment suffers a significant limitation, which is called “appropriate copy” problem. The appropriate copy of an article for one institution may be the version available from the publisher’s own web site, for another it may be the version held locally, and yet for another institution which does not subscribe to the electronic version of the journal in which the article is published, the appropriate copy may be the print copy on the library’s shelves (Walker, 2001). When a DOI link is “clicked”, the DOI is sent to the resolver, and the URL found in that database is returned to the browser as an HTTP redirect. The mechanism in its general form is therefore limited to supporting a one-to-one relationship between DOIs and URLs. This would be fine if there existed only a single copy of each electronic journal article. The limitation of one-URL-per-DOI was recognized as a significant issue from the beginning of the DOI implementation, and in fact the Handle System technology does allow multiple URLs to be associated with each DOI. However, even if multiple addresses were registered for a DOI, there is nothing in the current architecture of linking that could select among these multiple addresses to provide the appropriate one for a given user (Beit-Arie, 2001).

The SFX community and the DOI community are collaborating to integrate the OpenURL framework and the DOI framework, to achieve better results in pursuing their goals. DOI, DOI-X, CrossRef, Citation/Reference Linking,

OpenURL, SFX and MetaLib can provide complementary services offering valuable solutions for libraries. Lots of collaborating works has been done during the past few years demonstrates the importance placed by the stakeholders on finding ways to integrate the many and varied library resources in a meaningful way for library users.

The authors believe in limitation of discussion on technical issues which are addressed in this paper. Because information technology is changing in very fast speed, many technical advantages may create and disappear by new advantages. It is obvious that technology will continue to improve and we would see more technological developments in future with new features for electronic journals.

References

Baldwin, Christine, & David, Pullinger (2000). What readers value in academic journals. *Learned Publishing, 13*(4), 229-240.

Baron, Joel (1997). Why we need information identifiers. *Learned Publishing, 10*(2), 132-134.

Beit-Arie, Oren, & et al. (2001). Linking to the appropriate copy: Report of a DOI-based prototype. *D-Lib Magazine, 7*(9).

Caplan, Priscilla (2001a). A lesson in linking. *Library Journal and School Library Journal, 126*(17), 16-18.

Caplan, Priscilla (2001b). Reference linking for journal articles: Promise, progress, and perils. *Portal: Libraries and the Academy, 1*(3), 352-356.

Grogg, Jill E. (2002). Thinking about reference linking. *Searcher, 10*(4), 56-61.

Grogg, Jill E., & Carol, Tenopir (2000). Linking to full text in scholarly journals: Here a link, there a link, everywhere a Link. *Searcher, 8*(10), 36-45.

Hodgson, C. (2005). Understanding the OpenURL framework. *NISO Information Standards Quarterly, 17*(3), 1-4.

National Information Standards Organization (2005). *The OpenURL framework for context-sensitive services*. Retrieved September 16, 2006, from http://www.niso.org/standards/standard_detail.cfm?std_id=783

Sally, Morris (2006). *Getting started in electronic journal publishing* (5nd ed). Retrieved September 13, 2006, from http://www.inasp.info/pubs/epub/INASPepublishing.pdf

Van de Sompel, H., & Beit-Arie, O. (2001). Generalizing the OpenURL framework beyond references to scholarly works: The bison-futé model. *D-Lib Magazine, 7*(7/8). Retrieved September 20, 2006, from http://www.dlib.org/dlib/july01/vandesompel/07vandesompel.html

Walker, Jenny (2001). *Seminar on linking technologies*. Retrieved November 26, 2006, from http://edina.ed.ac.uk/projects/joinup/seminar.html