Refereed article

Open linking for libraries: the OpenURL framework

Jenny Walker

The author

Jenny Walker is with the Ex Libris Information Services Division, Watertown, Massachusetts, USA.

Keywords

Libraries, Library services, Information technology, Links, Electronic data

Abstract

Reports on the ground-breaking work by Herbert Van de Sompel while associated with Ghent University, Belgium, and at the Los Alamos National Laboratory, which has led to the creation of an open linking framework for scholarly research. The open linking framework proposed by Van de Sompel puts libraries firmly in control, allowing them to determine the scope of their institution's interlinked research environment both in terms of extent and range of link services offered to their users. It places librarians in the "trail-blazers" profession, predicted by Vannevar Bush in 1945: they enthuse in the task of establishing useful links between resources. The creation of such an open linking environment requires the collaboration of all the stakeholders in the information industry; and underlying the success of this is an emerging standard, the OpenURL, now on a fast track path to acceptance by NISO. Van de Sompel's research work led to the development of the SFX server technology, which was demonstrated in operation at both the Ghent University and the Los Alamos National Laboratory. Ex Libris subsequently purchased the SFX server technology from Ghent University in early 2000.

Electronic access

The research register for this journal is available at
http://www.mcbup.com/research_registers

The current issue and full text archive of this journal is available at
http://www.emerald-library.com/ft

New Library World
Volume 102 · Number 1163/1164 · 2001 · pp. 127–133
© MCB University Press · ISSN 0307-4803

Introduction

I first met Herbert Van de Sompel of Ghent University in Belgium in 1997 when, as SilverPlatter Technology Products Manager, I was assigned to work with Van de Sompel and his team to explore some linking proposals. The initial work focused on linking from citations in SilverPlatter A&I (abstracting and indexing) database records to the Ghent University OPAC to check local holdings information. Later, as e-journals proliferated, this work was extended to embrace linking to full text, and also further extended to include other types of links of interest to a researcher, such as links to citation databases. Ex Libris, as the supplier of the ALEPH 500 OPAC system at Ghent University, was also involved from the outset in these linking initiatives.

At the time, prevailing solutions for integrating an institution's heterogeneous information resources focused on uniform access to resources through Z39.50[1]. Van de Sompel was insistent that for his institution this was only a partial solution. While Z39.50 provides front-end integration for resource discovery it does not provide for the important back-end integration whereby a user, having chosen a specific resource in which to start their research, can link to other resources and services. Van de Sompel's perseverance has resulted in a ground-breaking framework of context-sensitive extended linking services for libraries that firmly puts the librarian in control to determine the manner in which the resources are interlinked.

Background

Libraries today buy licenses for an ever-increasing number of information resources from a range of different providers and use a diverse set of technologies for service delivery. In addition, relevant resources are freely available on the Web for libraries to incorporate into their e-collections and to make them readily available to their users. Materials may be in print and/or electronic form; formally and/or informally published; and stored locally, for access via an institution's intranet, or remotely and accessible via the Internet. A number of services are outside the library's control, but

New Library World
Volume 102 · Number 1163/1164 · 2001 · 127–133

nonetheless libraries want to integrate their resources, presenting the information from any particular source within the context of the complete collection. Searching across repositories is only part of the solution. Linking is also essential.

The simplicity and ubiquity of the Web is such that users now have high expectations of their own digital library environments in terms of ease of navigation between the many and varied electronic information resources. They also expect to spend a minimum of effort in their information-seeking activities.

It is the librarian, as information intermediary, who is very well positioned to determine how resources should be interlinked to provide for ease of navigation, i.e. what should be linked to what and how it should be linked. The librarian has traditionally applied a broader range of knowledge to packages of information and is now challenged in extending their skills to highly diverse and widely dispersed users wherever and whenever they need it.

Linking is not a new concept, though. In 1945, Vannevar Bush, in his ground-breaking paper, "As we may think", identified the importance of actionable links between conceptually related items (Bush, 1945). At that time he reported on the problems of filtering through the "summation of human experience [which] is being expanded at a prodigious rate". Information overload is not a new problem, but with the technology available to us today we are surely better equipped to manage this.

Bush described an approach he termed "associative indexing" and suggested that this process of selecting by association rather than through indexing might yet be mechanized. The Web has not yet shown itself flexible enough to support Bush's vision, to link everything to everything, under the control of the individual user. But breakthrough technologies are emerging that take a giant leap forward towards this; providing control at least at the level of the library.

Reference linking: a review

The Web is undoubtedly facilitating linking among digital items – and we see many excellent examples of this in current offerings from information providers – secondary and primary publishers, and from the aggregators – but linking must also take account of the large body of literature not yet available in digital form.

Citation reference linking is one important form of linking that has been in existence for a long time, although historically the linkage system is based on paper manifestations of the articles, not surprisingly called "papers". In a paper model, it was not necessary to distinguish between an intangible work such as an article and the printed manifestation of this – they were one and the same. However, in the new digital publishing environment the same article can exist, for example, in a print form, in PDF and/or HTML form, and it is often important to distinguish between these.

Today's linking systems in the electronic environment tend to have been built to mimic the paper paradigm. Linking of citations to the articles themselves often relies on identifiers from the paper paradigm – ISSN, Volume, Issue and Start Page number. However, there are some notable exceptions in the use of these identifiers. The American Physical Society (APS) stopped using page numbers in favour of article numbers for their e-journals. These are assigned at the time of posting a journal electronically and are consistent across the print and electronic versions of the article. Two years ago the *British Medical Journal* (*BMJ*) adopted their "Electronic Long, Print Short" or ELPS article policy that recognizes differing needs of practitioners and researchers. For certain articles the electronic journal, eBMJ, hosts the long version of an article for the researchers, while the print *BMJ* has a shorter version of the same article for the practitioners. The journal *Pediatrics* publishes some studies primarily in electronic form, with only an abstract in the paper version.

A shift is occurring from today where print tends to be authoritative and complete, to an era, now surely not far off, where electronic is considered authoritative and complete; and where business models do not consistently tie the purchase of electronic materials to the print purchase. Indeed, a number of publishers and journal aggregators have already started to offer pay-per-view options for articles.

Linking to date has also tended to be unidirectional, although there are some cases of bi-directional linking, or interlinking. Linking has also largely been actionable only to one level and does not allow, let alone

New Library World
Volume 102 · Number 1163/1164 · 2001 · 127–133

encourage, wider or deeper navigation for the user through the information landscape.

Van de Sompel recognized that linking systems that are in operation today tend to fall into two categories and are focused on that part of the linking framework in which links are constructed in order to link to something. Some solutions adopt a static linking approach to this aspect while others employ dynamic linking; yet others use a combination of both (Van de Sompel and Hochstenbach, 1999a).

Static linking

With static linking, the links are computed in advance using batch processes. In some cases, such as in journal publishing, the links may be established once, for all time. Some models provide for ongoing updates to the links such as SilverPlatter's SilverLinker solution, where a pre-computed database of journal article's links is accessed in real-time and actionable links are inserted into the record. Other models such as DOI/CrossRef and PubMed/PubRef, provide a central database of unique identifiers and associated metadata, plus a look-up facility whereby, for each set of metadata provided, an identifier is returned. The identifier is then inserted in a resource, as an actionable identifier which when "clicked" will point at a central resolver. There are many other examples of static linking.

Static links are advantageous in that the links will almost certainly lead to the desired target. However, all the required information for the creation of the links must be available to the creator of the links at the time of creation.

Dynamic linking

More recent linking initiatives have explored the opportunities presented by computing the links for an information entity when and only when required – "on the fly". This ensures that links are based on the most recent information. However, success depends on the quality of the metadata in the source document and cannot, therefore, be guaranteed. EBSCO's SoftLinks uses a dynamic linking solution for linking to full text; as does the Dialog@Site™ GOLD eDOCS solution. These use algorithms to dynamically generate URLs.

There is currently a lack of adopted standards to generate URLs for the dynamic resolution of links; so much must be known about metadata required for linking and about the link-to syntax for the target. Publishers increasingly have a standard syntax that allows a "deep" link to be created to lead directly to the article itself. Quite often, though, the lowest level to which a link can be made is to the journal home page or the journal issue table of contents. The user is then required to drill down further within the vendor's system to access the article itself.

Both the static and dynamic linking approaches just described can be characterised as being "closed"; the links are dictated by the authority offering the linking facilities and the local institution where the links are to be used has no, or few, means to act upon the link. Some of the solutions are, however, context-sensitive in that the local institution may have some control over the resolution of the links.

The "appropriate copy" problem

During 1999, two NISO/DLF/NFAIS/AAP sponsored Reference Linking workshops specifically set out to address the question: "Given a standard citation, how does one link to the thing to which that citation refers?" (Caplan and Arms, 1999b). However, the focus became limited to citations to journal articles; and, considering the fact that there may be many manifestations of a cited article, or work, an emphasis was placed on linking to the appropriate copy of the work. The previous question was reworded as "Given the information in a citation to a journal article, how does a user get from a citation to an appropriate copy of the article?" (Caplan and Arms, 1999b).

For example, a user accessing an electronic article from the Institute of Physics Publishing (IOPP) Electronic Journals Service may wish to navigate through the given information environment by linking from one of the articles shown in the reference list to the full text of the referenced article. The referenced article could be from an Elsevier Science journal that is available from Elsevier's own ScienceDirect Web-based service. Such a link is presented to the user by

New Library World
Volume 102 · Number 1163/1164 · 2001 · 127–133

IOPP through a CrossRef Link. However, there may be many manifestations of the article available, including a locally-held electronic copy such as that at OhioLink; and it is to this version of the article that the user should be linked. An individual user has, by virtue of their affiliation, rights to access a particular copy, the "appropriate" copy, and linking must take account of such user affiliation.

Multiple copies of a work may exist for a number of reasons. First, there is an issue of operational performance in that users will want to access objects that are "closer" to them in network terms. Second, different vendors provide different service models using the same content. The competition that then arises encourages functional innovation and rationalises the pricing structures. And finally, archiving is of extreme importance to libraries. Institutional failure could be as great a danger as technical failure, particularly when dealing with commercial players. Multiple copies of a work held by different parties are the best protection.

Extended linking services

While the delivery of full text to the desktop has been a major advance in recent years and offers a key service to researchers, the librarian should be able to define many additional services they consider relevant to their users.

In his research findings, Van de Sompel restated the Reference Linking problem to be "Given (any) bibliographic metadata, how does one present appropriate or 'context-sensitive' extended services?" (Van de Sompel and Hochstenbach, 1999c). In other words, linking from metadata – a full citation or a single identifier – should be to a wide range of services, and all kinds of links between electronic scholarly information resources should take into account the context of the user. These may include links to relevant local information repositories, service links to check the author, article and/or journal citation information; or a link to relevant Web-based resources, possibly via the emerging subject gateways. A link to the local OPAC, for example, may be of great importance where the user's institution does not subscribe to the electronic version of a particular journal, but does maintain a print subscription. In this case, the "appropriate copy" for the user is the one on the shelves in the library.

The OpenURL framework

As stated previously, in many of the established linking solutions used today, service links are presented to users in a manner that fails to take into account their context, i.e. the electronic library resources that are accessible to them. Further, many of these solutions also fail to offer any local control over what links are presented to users and how such links resolve. Such linking frameworks have been referred to as non context-sensitive and "closed". The lack of context in the provision of linking services causes serious problems, the appropriate copy problem being the most cited one.

Allowing institutional service components that do know about the context of the user, and that provide the local institution with control over the linking services they offer, may open such closed linking frameworks. These may be in addition to or instead of the services delivered by information providers.

In order to address the interoperability problems inherent in linking heterogeneous information resources, Van de Sompel introduced the open linking framework in which the fundamental concept is as follows: a resource introduces the OpenURL, a user clicks the OpenURL and the target of the OpenURL is a service component of the user's choice. Via the OpenURL, metadata is transported from the information resource to the user's chosen service component. If the user's service component is adequately tailored it can deliver context-sensitive services.

Van de Sompel proposed the creation of an overlay linking service, which is independent of the various information resources, and is under the control of someone of the user's choice, typically the library. Such overlay services must be flexible and easy to implement in order to maximize acceptance and minimize overhead. Further, there must be buy-in from the providers of the information services. Parallels exist here with subsequent work by Van de Sompel and colleagues on the Open Archives initiative (OAi) through which has developed

New Library World
Volume 102 · Number 1163/1164 · 2001 · 127–133

low-barrier metadata harvesting specifications.

By implementing the OpenURL, information providers have the opportunity fully to integrate their information resources into their customers' overall information services. By creating the hooks that link back to individual libraries, information providers allow libraries to add customized linking services based on the libraries' local resources and environment. As already mentioned, many companies in the information industry are embracing the opportunities created by the Web and are delivering interlinked services. Now, the OpenURL framework allows those companies to make a leap forward on this path, by enabling their customers to add local services that link to other institutional resources.

The OpenURL status

In December 2000, Van de Sompel and Oren Beit-Arie, Ex Libris, submitted the OpenURL to NISO for accreditation as an ANSI Standard; and it was accepted as a Fast Track work item. A growing number of information providers have developed or announced the ability to generate and output OpenURLs, and therefore their services will interoperate with an institutional service component. A list of these providers can be found at: http://www.sfxit.com/sources.html

The current draft specification of the OpenURL and other relevant material relating to the OpenURL can be found at: http://www.sfxit.com/OpenURL.html

The OpenURL and CrossRef

The CrossRef linking solution is gaining widespread acceptance and now has more than 50 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 Digital Object Identifier (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.

Based on an initial prototype by Van de Sompel, the US Digital Library Federation (DLF), CrossRef, the International DOI Foundation (IDF), the Corporation for National Research Initiatives (CNRI), and Ex Libris, are exploring concepts to address these CrossRef local resolution issues by integrating the OpenURL framework with the existing CrossRef/DOI infrastructure (NISO, 2000).

SFX: a practical example of an open linking overlay service

The initial findings from Van de Sompel's research at Ghent University, were extended in experiments which included the Los Alamos National Laboratory and primary publishers as participants (Van de Sompel and Hochstenbach, 1999c). Ex Libris (USA) Inc. acquired the resulting technology, the SFX server, from Ghent University in early 2000.

SFX is an OpenURL-compliant open linking solution for libraries. It is an institutional service component (or overlay service) that provides libraries with an independent means of offering seamless interconnectivity between their ever-increasing collections of information resources. The SFX solution allows links between these resources to be library-defined: reference librarians can choose appropriate content from a range of information vendors and interconnect these in a manner of their choosing. They can then provide service links that they feel are appropriate for their end-users. With SFX, they do not depend solely on the linking services defined by the information providers; or on a specific set of identifiers and/or communications protocols.

SFX for librarians

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. It also offers a single point for the administration of linking services; and for the collection of statistics. SFX provides a wealth of statistics on the

New Library World
Volume 102 · Number 1163/1164 · 2001 · 127–133

usage of distributed resources that has never before been available.

Within SFX, a database is maintained that describes an institution's collection. The librarian defines sources (e.g. Medline, Web of Science, Journal of Physics A), services (e.g. holdings, full text, table of contents, citation lookup, etc.) and targets (e.g. the library OPAC, Science Citation Indexes, the journal article full text repositories). "Thresholds"" then define what information is required for the link. The link source goes to the linking service to see what options are available, checks the thresholds and determines what services can be offered to a particular user.

Ex Libris provides a global database of resource templates that can be localised and customised by an institution to match their own collections. Tools are available within the SFX server that allow for the easy customisation of the global database and ongoing maintenance of local information. From the experience of a number of leading institutions that have implemented SFX it is clear that these tasks have relatively insignificant resource implications. Ex Libris can provide regular updates of the global database.

SFX for users

Users whose institutions have an SFX server will notice as they navigate resources – hosted locally by the institution or remotely by third parties – that an SFX button appears along with each retrieved reference. Clicking on the SFX button for a reference brings up an SFX menu window from which the user can select from a number of services that are relevant for that reference. These services are institution-specific and will lead users only to those resources and services deemed appropriate by their librarian. The librarian can determine the scope, definition and description of such services to be offered based on the information resource from which the user is linking, and on the content of the reference. SFX services will only be offered for a particular request if the target service is deemed to be able to resolve such a request. For example, a link to the electronic full text of an article will only be offered as a service to the user if:

- the article is known to be published in electronic form; and
- the user's institution has a valid subscription to the relevant journal issue.

But, of course, other relevant links can be offered, such as links to the local OPAC to check print holdings, or, subject to library policy, a link to a preferred document delivery service. The range of services offered to a user is limited only by the creativity of the librarian defining these services.

Conclusion

Arthur C. Clarke's Third Law from his book, *Profiles of the Future*, states that "Any sufficiently advanced technology is virtually indistinguishable from magic" (Clarke, 1962).

SFX, indeed, brings some of that "magic" to scholarly research. No matter where a user starts their search – in an OPAC, an A&I database or directly in an e-journal service – they will, wherever possible, be presented with SFX links that will lead them to relevant resources as defined by their librarians. No more "access denied" messages! With SFX, users can also experience the serendipity of searching for one item and, by association, being led to the discovery of other relevant and related material.

What seem to be proving most challenging in an interlinked environment are, not the technical challenges, but rather human and political barriers to interoperability. However, I am very pleased that the stakeholders in the information industry have largely embraced the OpenURL Open Linking Framework, and many of the information service providers have already implemented the mechanisms to enable this, or have committed to do so. This is very good news for libraries, and their users.

Note

1 Z39.50 is an international standard for communication between computer systems primarily, library and information-related systems. Z39.50 is becoming increasingly important to the future development and deployment of inter-linked library systems (http://www.biblio-tech.com/html/z39_50.html).

References and Web sites

American Physical Society (APS), http://publish.aps.org
British Medical Journal (*BMJ*), http://www.bmj.org
Bush, V. (1945), "As we may think", *Atlantic Monthly*, Available: http://www.isg.sfu.ca/~duchier/misc/vbush/vbush-all.shtml
Caplan, P. (1999a), "Report of the working group of the reference linking workshop". Available http://www.lib.uchicago.edu/Annex/pcaplan/reflink.html
Caplan, P. and Arms, W.Y. (1999b), "Reference linking for journal articles", *D-Lib Magazine*, Vol. 5 No. 7/8, Available: (D-Lib July/August 1999) http://www.dlib.org/dlib/july99/caplan/07caplan.html
Clarke, A.C. (1962), *Profiles of the Future*, Victor Gollancz, London.
Corporation for National Research Initiatives, http://www.cnri.reston.va.us/
CrossRef, www.crossref.org
The Dialog Corporation, http://www.dialog.com/
EBSCO Information Services, http://WWW.epmet.com
Elsevier Science (Elsevier), http://www.elsevier.com/
International DOI Foundation, http://www.doi.org
Institute of Physics Publishing, http://www.iop.org/EJ/welcome/
NISO (2000), "NISO/DLF/CrossRef Workshop on Localization in Reference Linking", Available: http://www.niso.org/CNRI-mtg.html
Open Archives Initiative, www.openarchives.org
Van de Sompel, H. and Hochstenbach, P. (1999a), "Reference linking in a hybrid library environment. Part 1: frameworks for linking", *D-Lib Magazine*, Vol. 5 No. 4, Available: http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html
Van de Sompel, H. and Hochstenbach, P. (1999b), "Reference linking in a hybrid library environment. Part 2: SFX, a generic linking solution", *D-Lib Magazine*, Vol. 5 No. 4, Available: http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt2.html
Van de Sompel, H, and Hochstenbach, P. (1999c), "Reference linking in a hybrid library environment. Part 3: Generalizing the SFX solution in the 'SFX@Ghent & SFX@LANL' experiment", *D-Lib Magazine*, Vol. 5 No. 10, Availabl: http://www.dlib.org/dlib/october99/van_de_sompel/10van_de_sompel.html
SilverPlatter Information Inc., www.silverplatter.com