The Role of Knowledge Bases in Improving Discoverability Now and in the Future – Why National and International Collaboration Is Key

Sarah Pearson
E-Resources & Serials Coordinator, Library Services, University of Birmingham and
Co-Chair KBART Working Group

Introduction

I am honoured to be able to talk to you at the CONCERT Conference this year on the subject of library technology now and in the future in enhancing the electronic library experience for our users. As an Electronic Resources Librarian at the University of Birmingham in the UK, I face many of the issues in managing e-resources on a day to day basis that are common to libraries across the world. I am also privileged to be involved with a project that has an impact on improving discoverability, maximising content usage, reducing knowledge base administration and developing best practice in our industry and it is this topic that I would like to discuss with you at CONCERT 2011.

Changing Expectations:

In less than a generation, we have shifted to a world where answers can be found at our fingertips. Today's 18 year old student has grown up with technology as part of their daily life. Accessing content electronically and engaging with people and discussing ideas over the internet is more comfortable to today's student than using the physical library. The whole of the information community from authors to editors, to publishers, to agents and intermediaries, and to libraries; have had to adapt to this generational shift in expectation and ways of consuming and producing information.

Today's academic has to read and digest much more information in order to make an impact and advance the field of study in which he or she operates. They have to produce more, which feeds more output into the scholarly communication supply chain which perpetuates the cycle. This is the reality in which libraries are delivering a service. There is simply more content available and libraries are often finding ways to provide institutional access to that content with less money and then find ways to manage access to that content with an often diminishing staff resource.

The Big Deal has been very important for academic libraries. It allows an institution to purchase a mass of content from a publisher with a comparatively low top-up fee to access any unsubscribed content. There are lots of negatives to the big deal which I won't explore in this presentation. However, in terms of access to scholarly output, it has revolutionised the industry. It has provided an appropriate response to the problem of funding access to content. However, the global economic environment has changed in the intervening decade and libraries are now trying to address how to continue affording the big deal and subscriptions generally without undermining their institutional teaching and research objectives.

With the wholesale shift in journal collections to online only, there are other factors such as digital preservation which have become of increasing importance. In the print arena, libraries have had a distinct role in stewardship and preservation of content in their physical library

buildings. In a digital environment, libraries have struggled to understand their role in the process of preservation as well as in access provision and enabling discoverability. The reality is that the librarian is no longer solely responsible for access and long term preservation of the content which they purchase for future generations. The whole community is responsible for curating and maintaining a record for future use. This is partly because of the restrictions placed in licences on local hosting of content leading to a shift in responsibility to the publisher 'guaranteeing' that a historical record of their output will be sustainable. This may be on the publisher platform but equally on 3rd party preservation services such as Portico and LOCKSS (Lots of Copies Keep Stuff Safe). In the UK, the expectation of long term access to publishing output and the changing global economic climate has led institutions to be even more interested in the level of access and sustainability of subscription content in case they need to cancel subscriptions to meet organisational budget cuts. We want to guarantee that we get and keep what we pay for on behalf of our users.In addition, National Libraries such as the British Library in London are keener than ever to ensure that historic record of electronic content is assured. So, the kinds of questions being asked in our community and in our relationships with content providers are:

CURRENT ACCCESS

- What are our access entitlements in terms of titles and coverage?
- How do we ensure that our users understand what they can access?
- How do they discover it on the wider Internet?
- Is usage of this content matching our expectations and investment?

CURATION AND PRESERVATION OF DIGITAL CONTENT

- What are the respective roles of the library, publisher, intermediary, author and national library in preservation of digital content?
- How do we evaluate and make use of preservation services such as Portico and LOCKSS?
- Are we confident enough of long term sustainability of access to go e-only?

Finally, the sheer scale of content being made available to our library users, means that automation and de-duplication of effort within institutions is key to providing a comprehensive and accurate experience for our combined customers. The benefits are not just for our users but reach back right through the supply chain. A great deal of effort is invested in making our library knowledge bases accurate and comprehensive but there remains much more work to be done. The duplication of effort by individual libraries is not sustainable and content linking and discoverability will suffer as a result. There is much more potential for library technology and industry standards to improve this provision and provide even more integration for our end users. This is where best practice initiatives such as KBART comes in. There are also new projects looking at improving efficiency in administration of e-resources and provision of 'above campus' solutions that will provide long term benefits to libraries locally, nationally and internationally. I will cover the shared service aspect in my presentation to you.

All this has led to a fundamental shift in thinking about integration and technology tools to support access and preservation

The Changing E-Resource Landscape:

The changing wider scholarly communication environment and external supply chain pressures as described above has informed a corresponding shift in the provision of access to content. The technology and applications have changed over the years to support more integrated access to a wide variety of different content types and formats. However, whether it be OPAC; link resolver; metasearch services; or vertical search resource discovery services; they all have one thing in common: They all rely on *metadata.*Whether it be the OPAC at one end of the spectrum or semantic web at the other, metadata is key to the presentation of online content and the ease of discoverability. The advent of the OpenURL Standard(NISO, 2005) and link resolvers which use OpenURL to improve linking to the 'appropriate copy' has vastly improved access and usage of online content in our community. Link resolvers have had a big impact on discoverability and content usage. From my own institutional perspective at University of Birmingham, we saw approximately a 40% increase in usage of full text journals the year after we implemented a link resolver. Now a mature product, link resolvers and their knowledge bases have also had a central role to play in the effectiveness of a new generation of resource discovery services such as Primo, Summon and EBSCO Discovery. Without the knowledge of which titles an institution subscribes to, such services wouldn't be nearly as effective. Institutions are already anecdotally reporting a marked increase in usage of content indexed in such services despite previous reports that usage seems to have been flattening out in recent years. It is on the topic of knowledge bases and e-resource metadata that I am honoured to speak to you at the CONCERT conference.

Why are knowledge bases so important to discoverability?

Before I talk about standards and best practice and their role in improving discoverability globally, I want to detail the problem statement that many e-resource librarians and users face daily.

A library may have access to a journal title through multiple access points:- via an aggregated database, publisher website, subscription agent gateway or institutional repository. A user affiliated with an institution may have access to the article through two of these access points. A link from the citation – perhaps in an abstracting and indexing (A&I) service – should direct the user to only those two access points: their 'appropriate copy'. Such individual configuration within each A&I database (or 'source') is difficult if not impossible to administer. In addition, library A-Z lists with links at journal level need to be manually maintained for each access point available to the institution. This level of manual intervention is unsustainable and may result in poor service delivery because staff resources are just not available to support it. In an OpenURL capable environment, a user comes across an article citation. It could be linked to the full text on a publisher's websiteor in a database or in a gateway or in a repository. But any one of these links might take the user to an "inappropriate" copy, i.e. one which they are not entitled to access.However, if the institution has a link resolver, they can register the base URL of that link resolver with the provider of the article.The provider also knows the metadata of the citation, and can put this together with the base URL of the link resolver to form a query.This query is directed to the link resolver, which contains a knowledge base of library and publisher holdings data. These are assessed to find matches i.e. the library indicates that it subscribes through that

provider, and the provider indicates how to link to its content.The link resolver can then put together a “regular” – predictable - link to the cited article to which the library has indicated that it has a licence. This means that libraries do not have to maintain each title in each access point individually but can do so in one place – via their link resolver which constructs a link spontaneously to the appropriate copy by combining metadata about the link resolver and the article citation. However, in order for OpenURL linking to be effective, the library must maintain a comprehensive, up to date and accurate knowledge base. Without this, OpenURL is useless.

The electronic resource supply chain has taken up the integration with link resolvers in terms of OpenURL compliance almost wholeheartedly over the last few years. However, the provision of accurate and up to date metadata to support the OpenURL creation via link resolver knowledge bases is still an issue for the supply chain to address. This is where the KBART project comes in. KBART (Knowledge Bases and Related Tools) is a joint UKSG and NISO project which was set up in response to a report commissioned by the UKSG and written by Scholarly Information Strategies (Culling, 2007). One of the recommendations of this report was that the adoption of a globally recognised best practice be implemented by content providers and link resolvers to make the supply of e-resource metadata to link resolvers more accurate, timely and comprehensive. The collaboration between UKSG in the UK and NISO (the National Information Standards Organisation) in the US aimed to deliver this best practice to the community. The project group consisted of publishers and content aggregators; link resolver vendors; and libraries and their consortia; as a representative group of the supply chain involved with improving the transmission of e-resource holdings. The aim was to supply a Recommended Practice to the community of common sense data fields that link resolvers required to create OpenURL links to their users. Alongside this, an Information Hub and series of educational and promotional activities were put in place to help inform and obtain requirements from the publishing and library community. In January 2010, the NISO Recommended Practice(NISO/UKSG Working Group, 2010) was published.

This Recommended Practice supports the aims of the original OpenURL standard in delivering total ‘link resolver compliance’. It specifies and describes the fields which content providers need to supply in order to present institutional users with access to their ‘appropriate copy’ as distinct from other targets on the Internet. It covers data fields like identifiers, title, coverage start and end dates of electronic holdings, embargoes and depth of coverage. It also recommends a file format and options for transmission to link resolver vendors. The Project has developed an endorsement program where content providers can test their metadata for compliance and celebrate their take-up of KBART through promotion of their compliance. A registry of supply chain contacts has been developed and content providers are encouraged to ensure that details of their metadata are discoverable on the KBART Registry (NISO/UKSG, 2011). Consortial licences are also starting to request KBART metadata in the publisher deal offering. For example JISC Collections (the national e-deal licensing and negotiation agency in the UK) are specifying KBART compliance in their 2012 licences. Phase II is now underway with the KBART group concentrating on refining the original guidelines to include more specific information relating to e-books and conference proceedings; consortial specific metadata; and open access holdings.

Knowledge bases and the future

So, this is where the community is currently. However, it is a very exciting time to be involved with library technology and access provision. Consortia around the world are starting to look at ways in which the library administration of e-resources can be made more efficient with the removal of the duplication of effort in link checking and knowledge base administration carried out by individual libraries on exactly the same content. For example in the UK, JISC Collections have been appointed by SCONUL and HEFCE as project lead on implementing a national knowledge base of metadata pertaining to access, entitlements, usage and holdings amongst other things. The intention of this knowledge base is to reduce the total cost of ownership to libraries in e-resource administration and for libraries to make use of this metadata in order to improve content linking without checking every title and access point individually. In addition, the service intends to facilitate use by commercial library technology vendors in importing this metadata to enhance the knowledge bases they provide to their customers. More information on the report which led to the commissioning of this project is available (SCONUL, 2011).

In the U.S. NISO are also exploring ways in which the provision of metadata to support Resource Discovery Services(NISO, 2011)based on central indexes can be standardised. With the growing maturity of such products and their success in improving the user experience still further, standardisation of metadata supply between and will ensure that the supply chain can continue to adopt integration with content providers and library technology vendors is becoming increasingly important. This will undoubtedly be much welcomed by the library community next generation library technology services to further improve discoverability and content usage.

In addition, JISC and the RLUK Resource Discovery Task Force in the U.K. have initiated a project called ‘Discovery’(Discovery, 2011) which aims to promote open metadata within the Information Community to ensure that metadata on collections can be re-used for the purpose of enhancing discoverability.

Finally (although I’m sure there are many more examples), NISO have implemented a review of the ERM landscape(NISO, 2011) to assess whether there are gaps in the provision of standards for the management of electronic resources. This builds on the work of the original Digital Library Federation ERM Initiative on which most commercial ERM systems have based their data fields and workflow.

In order for such initiatives and projects to be successful, they need to operate within the global scholarly information community with endorsement of all sectors of the community. There are no such things as international borders when it comes to supply of scholarly output. We need to integrate internationally as well as nationally, in order to provide a joined up and fully functional e-resource supply chain. Such efforts if truly global will deliver immeasurable benefits to libraries and their users. So, in conclusion, as to where I see electronic library provision in 10 years time, I would whole heartedly hope that a more integrated, intuitive, global and collaborative discovery environment would be firmly established. The information community needs to work together with the emerging tools and standards to ensure that this comes to fruition.

More detail on KBART can be found on the UKSG Information Hub and NISO website and in ‘Serials’.

References:

Culling, J. (2007) **Link Resolvers and the Serials Supply Chain: Final Project Report for UKSG** [online]. Available from: http://www.uksg.org/sites/uksg.org/files/uksg_link_resolvers_final_report.pdf [Accessed 10 October 2011]

Discovery (2011) **Welcome to Discovery: Towards a thriving metadata ecosystem** [online]. Available from: http://www.discovery.ac.uk [Accessed 10 October 2011]

JISC / SCONUL (2011) **Shared ERM Requirements Project** [online]. Available from: http://sconulerm.jiscinvolve.org/wp [Accessed 10 October 2011]

NISO (2005) **ANSI/NISO Z39.88-2004: The OpenURL Framework for Context-Sensitive Services** [online]. Bethesda: NISO Press. Available from: http://www.niso.org/apps/group_public/download.php/6640/The%20OpenURL%20Framework%20for%20Context-Sensitive%20Services.pdf [Accessed 10 October 2011]

NISO (2011) **ERM Data Standards and Best Practices Review** [online]. Available from: http://www.niso.org/workrooms/ermreview[Accessed 10 October 2011]

NISO (2011) **NISO Newsline October 2011** [online]. Available from: http://www.niso.org/publications/newsline/2011/wgconnectionoct11.html#new2 [Accessed 10 October 2011]

NISO/UKSG (2011) **KBART Registry** [online]. Available from: http://sites.google.com/site/kbartregistry [Accessed 10 October 2011]

NISO/UKSG KBART Working Group (2010) **A Recommended Practice of theNational Information Standards Organization (NISO) and UKSG** (NISO-RP-9-2010) [online]. Available from:www.niso.org/publications/rp/RP-2010-09.pdf[Accessed 10 October 2011]

Pearson, S. (2010) KBART – Improving content visibility through collaboration. **Serials** [online], 23 (2): 150-154. Available from: http://www.metapress.com/content/f12w477024115618/fulltext.pdf [Accessed 10 October 2011]