KBART – improving content visibility through collaboration

In recent years, link resolver technology has become integral to ensuring successful institutional access to electronic content. The corresponding take-up of OpenURL compliance amongst content providers in response has resulted in a global solution to the 'appropriate copy' problem. However, this solution is only effective if the knowledge base behind the link resolver is up to date, accurate and comprehensive and this is a factor that is often overlooked in establishing OpenURL compliance. This article explores the importance of OpenURL and knowledge bases to the information community as a whole and provides an overview and update of the role that the KBART (Knowledge Bases and Related Tools) project has to play in improving knowledge base metadata.

SARAH PEARSON
E-Resources & Serials
Co-ordinator
Library Services
University of Birmingham and
Co-Chair KBART group

The importance of link resolvers and their knowledge bases

The explosion of electronic content and the corresponding dilemma of how best to provide space for print collections in our libraries has led to a surge in the number of institutional subscriptions where electronic access is the main method of delivery. Such take-up provides well-documented benefits for users who can now access the content they need for study, teaching or research at any time of the day or night. It has also meant that a quiet revolution has occurred within libraries in the way that such access is managed. In an age where tens of thousands, if not hundreds of thousands, of individual electronic resources are delivered by the library service and are available through multiple access points, it is no longer an option to manage these links manually. Such efforts lead to the failure of links when URLs change and an inability to resource the effort needed to add links to new content, update existing links and remove those which are no longer active. In addition, the difficulty in linking users from an article citation to the version of the full text that is appropriate to their institution's subscription has led to much confusion and inefficiency. To address this 'appropriate copy' problem and (as an essential by-product) to reduce the administration performed and duplicated by libraries the world over, OpenURL and link resolvers were invented.

Firstly, what is OpenURL technology and what is a knowledge base? Figure 1 illustrates the linking scenario. An article could be available 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 or from Google Scholar – 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.

The OpenURL (Z39.88)[1] standard, developed by NISO in conjunction with its inventor, Herbert Van de Sompel and others, supplied a new schema for an OpenURL link to solve the 'appropriate copy' problem. An OpenURL is used to transfer
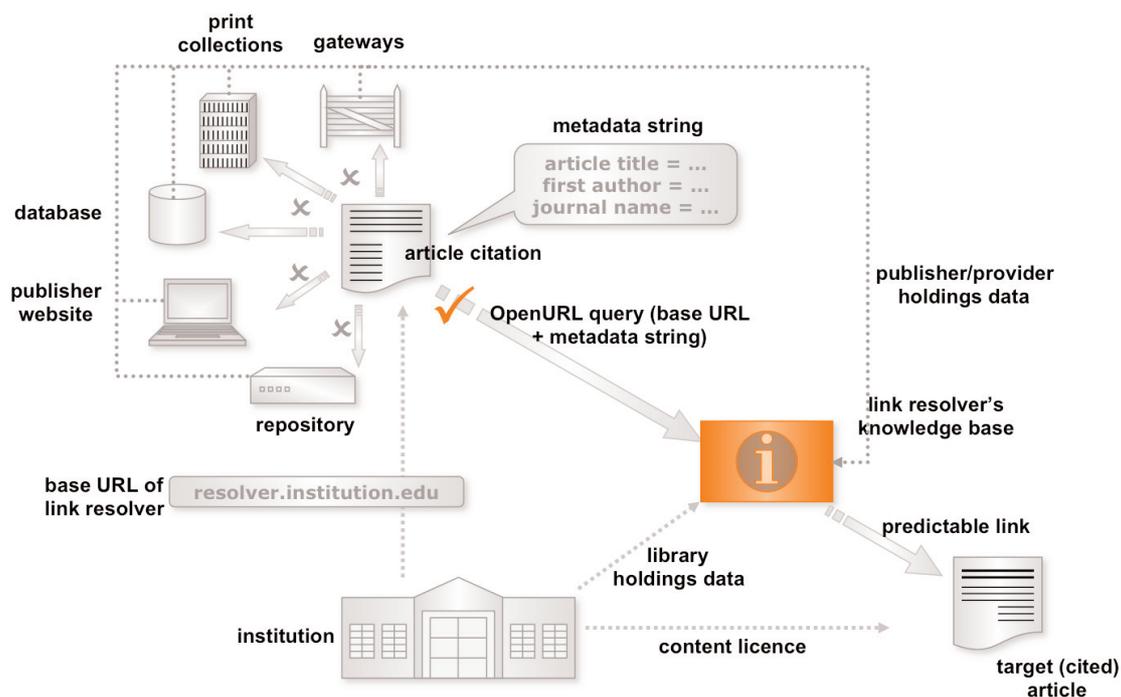
Figure 1. OpenURL technology enables linking to knowledge bases in a number of ways

metadata – describing the citation the user is coming from – to an institutional link resolver. The link resolver is populated with information about content that the institution has access to and delivers back to the user the appropriate links.

The OpenURL query itself contains two elements: the location of the link server for an institution (base URL) and the metadata string for that content (e.g. ISSN, journal title). The OpenURL link resolver is using its knowledge base to match these elements against available services and constructs the links to direct the user to the desired website.

This requires an institution to have a link resolver supported by a functioning knowledge base that provides information about full text and other appropriate services available to that user. The library sets up their link resolver 'base URL' as a one-time task with each content provider and then also 'localizes' the knowledge base so that the active links are based on their institutional entitlements.

For content providers, there are two elements to achieving 'link resolver compliance' – the adoption of the OpenURL standard and the provision of accurate and timely metadata to link resolver knowledge bases. The former is now widespread within the information community (although not without its problems[2]). However, it is the latter

that is often forgotten in achieving OpenURL compliance and still remains to be tackled effectively by the information community. The content provider is responsible for ensuring the timely, accurate and comprehensive delivery of global metadata describing their content that will then enable the link resolver vendor to populate the knowledge base with this data and the library to select the 'appropriate copy' for their users from the knowledge base. Ultimately, OpenURL compliance is no good if a library cannot select the appropriate service in the knowledge base for their users. The efficiency and effectiveness of OpenURL linking requires the involvement of all parts of the supply chain, from primary publishers, hosting services and aggregators, through to link resolver vendors and libraries.

Why is this important? It matters to the entire supply chain for electronic content because knowledge bases and OpenURL enable content linking to be predictable and efficient regardless of different institutional entitlements. By increasing visibility and accuracy of links, users are more likely to find the content and access it successfully, resulting in greater usage of electronic content. Any positive impact on content usage is more likely to be translated into a favourable renewal decision by the institution and consequent protection of revenue for content providers. At a time

when value for money and demonstration of that value is crucial, exposing holdings metadata to link resolver knowledge bases alongside OpenURL compliance is important for all in the supply chain – those publishing, working with, supplying, administering and funding electronic content.

Link resolver knowledge bases have become a highly valued tool to libraries. They establish what the users from that institution can access, both in terms of titles and of coverage years within that title. As the metadata within knowledge bases has grown over time, so has the emphasis placed on maintaining the accuracy and comprehensiveness of that metadata. Libraries invest very heavily in keeping this knowledge base up to date as do the link resolver vendors themselves. With the complexity of consortial and institutional entitlements, the plethora of content packages and the emerging complexity associated with hybrid and open access content, this is no easy task. It is being recognized that the whole supply chain has a role to play in keeping knowledge bases accurate. In fact, the best way to achieve this is at source, by content providers providing regular updates of holdings metadata to knowledge bases.

Knowledge bases are also becoming more important over time as a collection development tool. With the huge amount of information on content provider holdings and the level of customization, link resolvers are uniquely placed to offer tools such as overlap and gap analysis between current collections and potential acquisitions. It is therefore becoming an important tool for reaching the library market and consequently inaccurate metadata can be detrimental to all stakeholders.

Although the benefits start with the end-user, they reach back right through the supply chain. However, despite a great deal of effort put in to keeping knowledge bases accurate, there remains much work to be done. The duplication of effort required by many libraries is not sustainable and content linking will suffer. A lot of the work currently undertaken by libraries is reactive – responding to problems that have been reported. A far better scenario is to proactively feed in changes to content metadata at the right time to enable timely links to be set up to content. By establishing regular, frequent exchanges of metadata between content providers and link resolvers, this effort can be minimized and the investment in good quality content linking assured. Without good practice recommendations and ownership of the problem

within the supply chain, users will continue to be confused by dead links, out-of-date holdings information and lack of access to content to which they are entitled. This is where KBART comes in.

The role of KBART in improving knowledge base metadata

These issues are central to the work of the KBART group, an initiative of NISO (the US National Information Standards Organization) and UKSG (a non-profit organization that connects the information community) that is exploring data problems within the OpenURL supply chain.

The KBART working group was set up in January 2008 by the NISO and UKSG management committees with two chairs appointed: Peter McCracken (for NISO) and Charlie Rapple (for UKSG). Its work is governed by the NISO Discovery to Delivery Topic Committee and the UKSG Main Committee. This collaboration ensured that any resulting recommendations would have global reach to match the availability of publishing output across borders. The work of KBART responds to recommendations in a research report commissioned by UKSG in 2007 and written by James Culling of Scholarly Information Strategies. The report entitled 'Link resolvers and the serials supply chain'[3] recommended that a 'code of practice' be produced on the methods and frequency of metadata transfer, along with the metadata elements required. Alongside this, education, promotion and communication activities should be considered in order to promote adherence to the code of practice.

The KBART working group consists of publishers, other content providers, link resolver suppliers and libraries with expertise in the field of knowledge base metadata. The main areas of activity were established as being best practice guidelines, educational materials and events, and a web hub to act as a central resource for knowledge base information.

Alongside the main report and guidelines, a series of other documents are available on the UKSG website. These include a glossary of terms, FAQs on OpenURL and knowledge bases, a description of supply chain roles and responsibilities for metadata transfer and an entry-level description of OpenURL technology.

In terms of the 'Recommended Practice' itself, the various problems resulting from poor metadata in knowledge bases were analysed, with input from the information community at various events over the last two years. A list of priorities to be addressed was then developed:

- identifier inconsistencies
- title inconsistencies
- incorrect date coverage
- inconsistent date formatting
- inconsistencies in content coverage description
- embargo inconsistencies
- data format and exchange
- outdated holdings data
- lack of customization.

Many of the priorities above are addressed in the phase I recommendations which were released in January 2010. However, after much debate, the KBART working group took the decision to focus on the more fundamental metadata issues, such as accuracy of global coverage description, before concentrating on areas like consortial- and institution-specific metadata. This approach was tested with content providers prior to release of the Recommended Practice. This was highly valuable in demonstrating that even the core metadata, such as coverage and identifier information, is difficult to provide accurately and uniformly. The resulting list of data elements within the KBART Recommended Practice is described in Table 1.

The recommendations also include the method and frequency of exchange, the data file format and naming convention along with descriptions of the data field requirements. They are designed to be used both by content providers who have never sent metadata to link resolvers before and by content providers who are already doing so but can use this as a checklist to review their current procedures.

Next steps

A number of next steps have been identified in the KBART Recommended Practice. These include reviewing requirements for consortial- and institution-specific metadata and also non-journal metadata. There has been considerable debate on mailing lists regarding the knowledge base metadata problems associated with open access, hybrid and free content. These are increasingly being activated by libraries alongside subscription content in order to provide a more comprehensive knowledge base for users. KBART will be reviewing the requirements for best practice in this area.

As recently publicized in a UKSG press release[4], the KBART group has established an endorsement framework around the Recommended Practice. This is designed to allow content providers and link resolvers to transition to the KBART format for e-holdings metadata and publicize and celebrate adoption of good practice. There is also a contacts registry where supply chain members can provide contact details and links to metadata files. More information on endorsement can be found on the KBART Information Hub[5].

Header	Description
publication_title	Publication title
print_identifier	Print-format identifier (i.e., ISSN, ISBN, etc.)
online_identifier	Online-format identifier (i.e, eISSN, eISBN, etc.)
date_first_issue_online	Date of first issue available online
num_first_vol_online	Number of first volume available online
num_first_issue_online	Number of first issue available online
date_last_issue_online	Date of last issue available online (or blank, if coverage is to present)
num_last_vol_online	Number of last volume available online (or blank, if coverage is to present)
num_last_issue_online	Number of last issue available online (or blank, if coverage is to present)
title_url	Title-level URL
first_author	First author (for monographs)
title_id	Title ID
embargo_info	Embargo information
coverage_depth	Coverage depth (e.g., abstracts or full text)
coverage_notes	Coverage notes
publisher_name	Publisher name (if not given in the file's title)

Table 1. Data elements within the KBART Recommended Practice

We would now urge all in the information community to review the KBART Recommended Practice and prioritize take-up of the guidelines within their organizations[6].

References

1. OpenURL Standard
   www.niso.org/standards/z39-88-2004 (accessed 9 June 2010).

2. OpenURL Quality Metrics Project
   http://www.niso.org/workrooms/openurlquality (accessed 7 June 2010).

3. Link Resolvers and the Serials Supply Chain: Final Report
   http://www.uksg.org/projects/linkfinal (accessed 7 June 2010).

4. KBART Press Release 1 June 2010
   http://www.uksg.org/news/kbartmay10 (accessed 9 June 2010).

5. UKSG KBART Information Hub
   http://www.uksg.org/kbart (accessed 7 June 2010).

6. KBART Recommended Practice
   www.niso.org/publications/rp/RP-2010-09.pdf (accessed 7 June 2010).

Article © Sarah Pearson

■ **Sarah Pearson**
E-Resources & Serials Co-ordinator
Library Services
University of Birmingham
Edgbaston
Birmingham B15 2TT, UK
Tel: +44 (0)121 414 5825
E-mail: s.pearson.1@bham.ac.uk