



Link resolvers and the serials supply chain: a research project sponsored by UKSG

In 2006, a UKSG-funded research study explored the data supply chain that has developed in recent years to facilitate the creation of link resolver knowledge bases. Through a combination of interviews and an online survey, a sample of content providers, subscription agents, link resolver suppliers, librarians and other stakeholders was consulted, to better understand this supply chain and any challenges it faces. The final report documents the current data flow arrangements and the roles, relationships and expectations of participating stakeholders. In addition, a number of issues and barriers to improving and extending the data flow for the benefit of library users are identified. Two central recommendations are made. Firstly, a Project COUNTER-style initiative should be established to define a code of practice for effective knowledge base supply chain participation. Secondly, stakeholders should build on early experiments with XML data formats and web services technology to accelerate and automate data transfers.



JAMES CULLING

Online Project Manager
Oxford University Press

Industry context

The OpenURL¹ provides a means for librarians, via link resolvers, to take charge of directing users at their institutions or organizations to appropriate, subscribed resources for content, whether in electronic or print form. OpenURL linking not only improves the online working environment for library patrons by reducing the number of linking dead ends but it also – by improving content visibility – increases the usage of the library's licensed and subscribed materials and potentially reduces document delivery spend. All of these are appealing outcomes for librarians.

However, the OpenURL is only one of two key components that make a context-sensitive linking framework possible. The knowledge base that underpins the link resolver is also a critical piece of this framework.

Currently, a number of commercial link resolver suppliers independently collect and collate data regarding the different incarnations of online journal and book content from many different information providers, to create proprietary knowledge bases for their own products. Some libraries have also built knowledge bases themselves as

part of 'home-grown' resolver applications, such as Gold Rush².

In this 'distributed' knowledge base model, and in the competitive marketplace for link resolvers, the knowledge base is used as a key differentiator in sales discussions. Its accuracy and comprehensiveness plus the frequency with which it is updated are all arguments used to persuade the librarian to opt for one solution rather than another.

Having selected a linking solution from the marketplace, the librarian (or their resolver supplier) localizes the 'out of the box' knowledge base to reflect local subscriptions and conditions: subscribed or preferred web resources, content packages and individual journal/book titles are made active in the system. Additional resources such as the local library catalogue are also configured to ensure that links to print materials are offered to users where relevant.

A link resolver then draws on its configured knowledge base for a given institution to determine the appropriate link(s) to offer to a user for a specific OpenURL. Whilst the OpenURL is the enabling technology that provides the link resolver

with key input data, it is the interaction with the knowledge base that – critically – determines the appropriate options for a particular citation or reference and delivers the user to their chosen destination.

Project context

As a result of the significant value they add and the local control over linking they provide, link resolvers have risen rapidly in profile in a short space of time and are now viewed by many academic librarians as an essential software component in their technology toolkit. It is therefore imperative, as digital collections become more and more critical to libraries, that the data residing in link resolver knowledge bases is current, accurate and reliable, if users are to discover and access the content that is selected and acquired for them by librarians.

And yet the experience of many librarians is that whilst resolver technology has a real potential to enhance access to digital collections, in practice it has also introduced a range of new problems: there can be significant delays in the updating of knowledge bases; the information about titles in packages from content aggregators can be inaccurate; and identifying who needs to do what to solve such problems can be difficult.^{3,4}

In July 2006 the UKSG invited tenders for a research project to explore the new data flow, or supply chain, that has developed to facilitate the creation of knowledge bases by resolver suppliers. This supply chain involves a number of organizations: publishers and other content hosts, subscription agents, librarians, providers of link resolver software tools, and others.

By exploring the views of the various parties it was felt that a study would be able to better understand the present supply chain, clarify current roles and expectations, identify any performance issues and barriers that need to be overcome to ensure a smooth flow of data to the end-user, and consider how the problems identified in the supply chain might be alleviated.

Scholarly Information Strategies (SIS) began work on the project in September 2006, with the final project report submitted to UKSG at the end of January 2007.

Methodology

SIS undertook 30 one-to-one telephone or face-to-face interviews with representatives from publishers, content hosts, link resolvers, subscription agents, libraries and other stakeholder organizations in the knowledge base supply chain. SIS sought to identify contacts for interview that would ensure a spread across organization size, geographic location, and – in the case of libraries – the resolver system deployed. The link resolver suppliers interviewed were EBSCO (LinkSource), Ex Libris (SFX), Innovative Interfaces (WebBridge), OCLC Openly Informatics (1cate), Serials Solutions (Article Linker) and TDNet (TOURresolver).

In addition, SIS sent a request for written feedback to the study's main questions to librarians via the major library list servs (liblicence-l, lis-e-journals and lis-serials).

Organizing interviews with librarians outside the UK and USA proved challenging. However, with the emergence of a number of recurring themes from the interviews and list serv feedback, SIS formulated a short online survey that addressed specific topics. Invitations to complete the survey were sent to library contacts drawn from the company's own survey database. No e-mail invitations were sent to UK contacts in the database as it was felt that sufficient feedback from this market had been captured by other means.

Of the 118 librarians who completed the survey, 90% worked in an academic institution (with or without a research programme). In terms of geographic breakdown, 54% of respondents were from North America, 26% were from Western Europe (not including the UK), 8% were from Australasia, and a further 8% were from Eastern Europe.

Results

In its final report to the UKSG, SIS grouped the findings of the study into a number of topics. These topics equate to the main objectives of the work (outlined above). Summaries of these same topics are provided below, but readers should see the final report for full details.⁵

Description of the supply chain

The final report provides both a knowledge base data flow diagram (a generalized schema of how information moves between stakeholders in the supply chain) as well as a roles/relationships matrix. The data flow diagram is reproduced in Figure 1.

Furthermore, the report discusses a number of complicating factors that impact the knowledge base data flow, as well as data interactions with CrossRef and Google Scholar.

The current knowledge base data supply chain is characterized by a complex series of roles, relationships and inter-dependencies between stakeholders. The major characteristics are:

- a number of link resolver suppliers creating or sourcing knowledge base data from publishers, other content hosts and subscription agents for their own proprietary systems (i.e. a 'distributed' model). The accuracy, comprehensiveness and currency of knowledge base data is a source of competition between suppliers.
- the reliance of resolver suppliers on data from content providers to populate their knowledge bases. This data is of varying quality, and its quality may or may not be improved by individual resolver suppliers prior to its delivery to libraries.
- a dependency by libraries on the data in knowledge bases (including the holdings details they source from content providers and subscription agents) for accurate and reliable linking provision to their users.

Issues and barriers

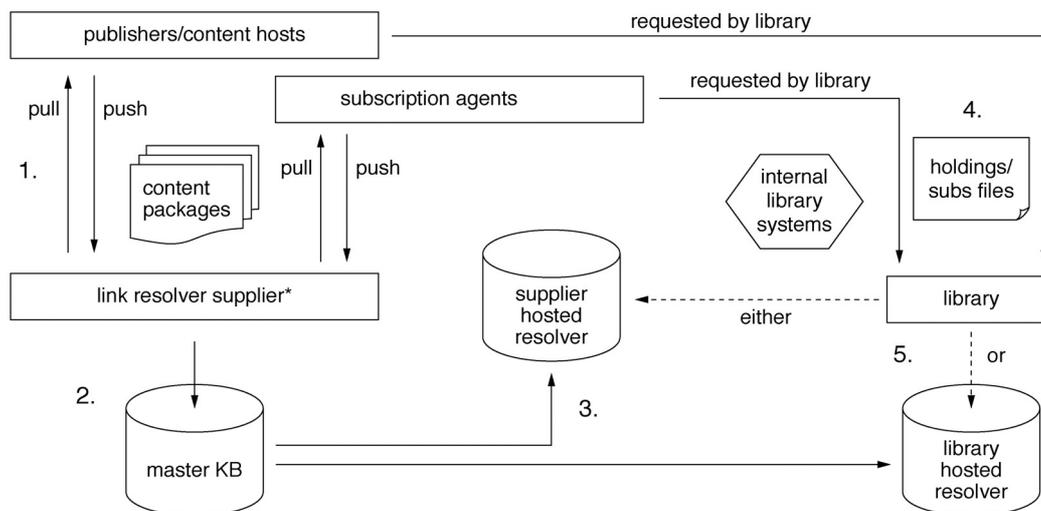
The study identified 12 separate issues/barriers in the present supply chain and the final report discusses these in full:

- lack of awareness
- lack of co-operation
- inaccurate and incomplete data
- content package issues
- journal title changes and transfers between publishers
- who has responsibility for data quality
- lack of data standards
- timing issues
- inbound linking issues
- OpenURL issues
- the role of the subscription agent
- the need to broaden knowledge bases.

Despite the existence of commercial link resolver services since 2001, the major barriers to improving the current situation for libraries and users are still a lack of understanding of this technology by stakeholders and a lack of closer co-operation between stakeholders.

Whilst some content providers are very aware of the role of link resolvers and the significance of data feeds to them for driving traffic to their content, there remains a significant number that do not make their collection details available to resolver suppliers at all, simply through not realizing that this is a desirable thing to do.

Conversely, whilst link resolver suppliers state that the level of co-operation from some content



* There are a number of resolver suppliers building their own proprietary Knowledge Bases (KBs) for the market

Figure 1. Generalized description of knowledge base data flow

providers is still not all that it might be, many publishers comment that a lack of open engagement and transparency regarding knowledge base requirements from the link resolver suppliers (as a group) has been problematic for them.

Other major findings:

- where data is provided to link resolver suppliers and libraries by content providers, a lack of understanding or appreciation as to the use to which this data will be put may be a factor in its incompleteness and inaccuracy
- most of the link resolver suppliers have separately invested much time and staff resource in working around difficulties with data from content providers, rather than trying to address the problems at source
- there is a lack of clarity and transparency in the supply chain regarding standards for data formats, expected frequency of data updates, construction of inbound linking syntaxes and OpenURL support
- most organizations that supply data in a structured format to link resolver suppliers and libraries today do so using comma- or tab-separated files.

However, there is no consistency in format and field labelling from one provider to another, creating manual effort for the library, and a huge data normalization problem that is duplicated across each of the link resolver suppliers.

- competition between organizations in the supply chain can limit co-operation and data sharing.

Many of the issues identified hinder broader adoption and limit the pace of information transfer through the supply chain, restricting the potential of link resolver systems.

The study concludes that whilst the community's attention in this area has been mostly focused on what it means to be OpenURL compliant, a code of practice and information standards to ensure optimal knowledge base compliance by all relevant parties have been sorely absent and overlooked.

Recommendations

The final report proposes two key areas of action to address the issues and barriers identified in the supply chain.

Firstly, that the education, communication and lack of transparency issues could be most effectively addressed via a mirror organization to that which operates in the usage statistics space, i.e. Project COUNTER⁶. This organization would seek to bring stakeholders together to define a visible code of practice for effective participation in the knowledge base supply chain. Librarians could then point content providers, subscription agents and link resolver suppliers towards these guidelines and, ultimately, require compliance via content and software licensing agreements.

A number of the content providers interviewed voiced their support for such a visible benchmark, arguing that whilst it would perhaps result in work for them in the short term it would reassure them that the effort they were investing was appropriate, worthwhile and valued.

Publisher, library and intermediary representatives should all be involved in this initiative in order to foster an open, inclusive and interactive discussion, in a manner similar to Project COUNTER.

The final report provides the parameters and probable values for four major areas of recommendation within a code of practice, as a starting point for further debate between stakeholders.

Secondly, the study identified that there are significant opportunities for stakeholders to explore the acceleration and automation of knowledge base data transfer, addressing data normalization and timing issues further.

A considerable number of interviewees, but especially the link resolver suppliers, see potential in the ONIX Serials Online Holdings (SOH) XML format⁷ as a common data format for resolver knowledge base information – both the global collections of content providers and the library-specific, local holdings details that are used in localization.

With one common data format for all exchanges, the normalization effort for the resolver suppliers would be reduced to one processing script for data from any source. This is clearly very appealing and would enable much more rapid data processing and more frequent knowledge base updates.

To date, progress with deploying ONIX SOH for this purpose has, however, been slow. There are a number of possible explanations for this, including perhaps a general lack of will to begin using the format for this purpose. However, a number of parties emphasized in discussion the potential for

problems arising from the size and complexity of ONIX SOH messages (especially for the larger publishers with many titles in their collections).

The link resolver suppliers and some content providers also recognize the potential of XML-based web services for automating (and therefore speeding significantly) data movement around the supply chain⁸. For example, RSS could be utilized as a mechanism to alert link resolver systems to changes in publisher collections (titles lost/titles gained, coverage of a title expanded/reduced, etc.), and possibly to initiate the retrieval of revised collection files. In addition, and again drawing on a comparison to the usage statistics arena, a SUSHI⁹ equivalent – a standard XML-based protocol for machine-to-machine harvesting of ONIX SOH data files – has potential for dramatically reducing the current drag in the timing of knowledge base updates.

There is a clear need for further experimentation with web services technology in this area, perhaps in conjunction with a stripped-down ONIX SOH format. The unique information about each library's holdings is the harder information to source and load (whether it is the library or its resolver supplier doing the work). Therefore, experiments between link resolver suppliers and both content providers and subscription agents should focus here. Whilst subscription agents have data protection and competition concerns arising from a closer collaboration with link resolver suppliers (some of whom have subscription agent divisions within or associated to their businesses), the study concludes that one of the greatest opportunities in the existing supply chain is greater co-operation between these two stakeholders.

References

1. ANSI/NISO Z39.88 -2004, The OpenURL Framework for Context- Sensitive Services, National Information Standards Organization (NISO): http://www.niso.org/standards/standard_detail.cfm?std_id=783
2. Gold Rush – A Discovery and Management Tool for Electronic Resources: <http://grweb.coalition.org/>
3. Chen, X, Assessment of Full-Text Sources Used by Serials Management Systems, OpenURL Link Resolvers, and Imported E-Journal MARC Records, *Online Information Review*, 2004, 28(6), 428–434.
4. Wakimoto, J C, Walker, D S, and Dabbour, K S, The Myths and Realities of SFX in Academic Libraries, *The Journal of Academic Librarianship*, 2006, 32(2), 127–136.
5. *Link Resolvers and the Serials Supply Chain*, Final Project Report for UKSG, Authored by James Culling of Scholarly Information Strategies, January 2007: <http://www.uksg.org/projects/linkfinal>
6. COUNTER – Counting Online Usage of Networked Electronic Resources: <http://www.projectcounter.org/>
7. ONIX for Serials, SOH: Serials Online Holdings – Release 1.0 (revised September 2005), NISO/EDItEUR: http://www.editeur.org/onixserials/ONIX_SOH1.0.html
8. EBSCO and OCLC Successfully Deploy Web Service for ONIX for Serials Online Holdings (SOH) Format – Technology Will Automate E-Holdings Delivery and Access, *EBSCO Support News*, June 2006: http://support.epnet.com/support_news/detail.php?id=250&t=h
9. NISO Standardized Usage Statistics Harvesting Initiative (SUSHI), National Information Standards Organization: http://www.niso.org/committees/SUSHI/SUSHI_comm.html

Explanatory note

Scholarly Information Strategies (SIS), the consultancy that undertook this research project on behalf of UKSG, ceased trading at the end of January 2007. James Culling is now working in a different capacity for Oxford University Press.

■ James Culling
 Online Project Manager
 Oxford University Press
 Great Clarendon Street
 Oxford OX2 6DP, UK
 Tel: +44 (0)1865 354851
 Email: james.culling@oup.com

For a link to the final Project Report on Link resolvers and the serials supply chain, published openly on the UKSG website, click here:

<http://www.uksg.org/projects/linkfinal>

To view the original copy of this article, published in *Serials*, the journal of the UKSG, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=20&issue=2&spage=111>

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=20&issue=2>