



# Link Resolvers and the Serials Supply Chain

*Final Project Report for UKSG*

A large, light grey, stylized graphic of the letters 'SIS' is positioned on the left side of the page. The letters are bold and sans-serif, with a slight shadow effect. The 'S' is on the left, the 'I' is in the middle, and the 'S' is on the right. The graphic is partially overlaid by a white, rounded shape that looks like a page corner being turned up.

**SIS**

Scholarly Information Strategies  
Oxford Centre for Innovation  
Mill Street  
Oxford  
OX2 0JX  
England  
Tel: +44 (0)1865 812058  
Fax: +44 (0)1865 793165  
Email: [info@scholinfo.com](mailto:info@scholinfo.com)  
Web: [www.scholinfo.com](http://www.scholinfo.com)

*By James Culling*

## TABLE OF CONTENTS

|   |    |
|---|----|
| Acknowledgements.....                                   | 3  |
| 1. Summary of Findings.....                             | 4  |
| 2. Industry Context.....                                | 7  |
| 3. Project Context.....                                 | 12 |
| 4. Methodology .....                                    | 14 |
| 5. Description of the Supply Chain .....                | 19 |
| 6. Issues and Barriers .....                            | 27 |
| 7. Recommendations.....                                 | 36 |
| 8. A Centralised Alternative.....                       | 43 |
| 9. Areas for Further Research.....                      | 47 |
| 10. References .....                                    | 48 |
| Appendix 1: Invitation to Interviewees .....            | 50 |
| Appendix 2: Invitation to List Servs .....              | 51 |
| Appendix 3: Text Copy of Online Survey .....            | 52 |
| Appendix 4: About Scholarly Information Strategies..... | 59 |

## **ACKNOWLEDGEMENTS**

Scholarly Information Strategies (SIS) would like to extend its gratitude to all those who freed up their valuable time to discuss this project, whether via a face-to-face or telephone interview, a written response to a request for feedback, or by completing the online survey.

Our thanks must also go to those who kindly agreed to review and comment on the contents of the final report.

## 1. SUMMARY OF FINDINGS

### Description of the Supply Chain

- The current knowledge base data supply chain is characterised by a complex series of roles, relationships and inter-dependencies between publishers, other content hosts, subscription agents, link resolver suppliers, libraries and others.
- The major characteristics are:
  - A number of link resolver suppliers creating or sourcing knowledge base data 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

- Despite the existence of commercial link resolver services since 2001, the major barriers to improving the current situation further for the benefit of libraries and users are a lack of understanding by stakeholders and a lack of closer co-operation between them.
- 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 realising that this is a desirable thing to do.
- Whilst link resolver suppliers state that the level of co-operation from some publishers 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.
- 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 the data will be put may be a factor in 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. Many have concluded that full text aggregators in particular focus their energies in other areas and metadata accuracy is never (voluntarily at least) going to be of high concern to them.

- Competition between organisations in the supply chain sometimes hinders co-operation and data sharing.
- 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. These issues hinder broader adoption and limit the pace of information transfer through the supply chain, restricting the potential of link resolver systems.
- Whilst the community's attention 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* have been sorely absent and overlooked.

### **Recommendations**

- The major issues in the supply chain could be most effectively addressed through education and communication via a mirror organisation to that which operates in the usage statistics space (Project COUNTER). This organisation, which the UKSG could take the lead in establishing, would seek to bring stakeholders together to define a visible code of practice for effective participation in the knowledge base supply chain. Libraries could then point content providers, subscription agents and link resolver suppliers towards these guidelines and – ultimately – require compliance via content and software licensing agreements.
- This study provides the parameters and probable values for four major areas of recommendation within a code of practice, as a starting point for further discussion.
- At the same time as addressing short-term needs via education and a code of practice, stakeholders in the supply chain should begin to explore the use of web services technology, most likely in conjunction with a stripped-down ONIX Serials Online Holdings (SOH) XML file format, for accelerating and automating data transfer.
- One of the greatest opportunities in the existing supply chain is further (automated) co-operation between link resolver suppliers and subscription agents. Through assisting the library in the knowledge base localisation task more directly, the subscription agent could play a very valuable role.

### **A Centralised Alternative**

- Many content providers and librarians are attracted to a more centralised model of knowledge base creation and management. They foresee communication and visibility benefits arising from a central system.
- Link resolver suppliers envisage difficulties in the operation of a completely centralised solution, and emphasise the significant task of data quality assurance (a task they currently undertake themselves). There is little enthusiasm amongst the link resolver suppliers for giving up this function, and certainly not whilst processes and standards to address data quality/consistency/comprehensiveness at source are still absent from the supply chain.

- A centralised solution would require investment from some (as yet unknown) source or a charging model to operate successfully, and the quality assurance/management costs are likely to be substantial. Such a model would effectively be in direct competition with the existing link resolver suppliers and does not appear viable in the short-to-medium term.
- Without a quality assurance layer, a centralised approach becomes more viable but the appeal for libraries is reduced.
- The CrossRef organisation is interested in becoming a channel for the distribution of its members' knowledge base data to all of the existing link resolver systems. This is a potentially viable and valuable addition to the supply chain, worthy of exploration in the short-to-medium term.

## 2. INDUSTRY CONTEXT

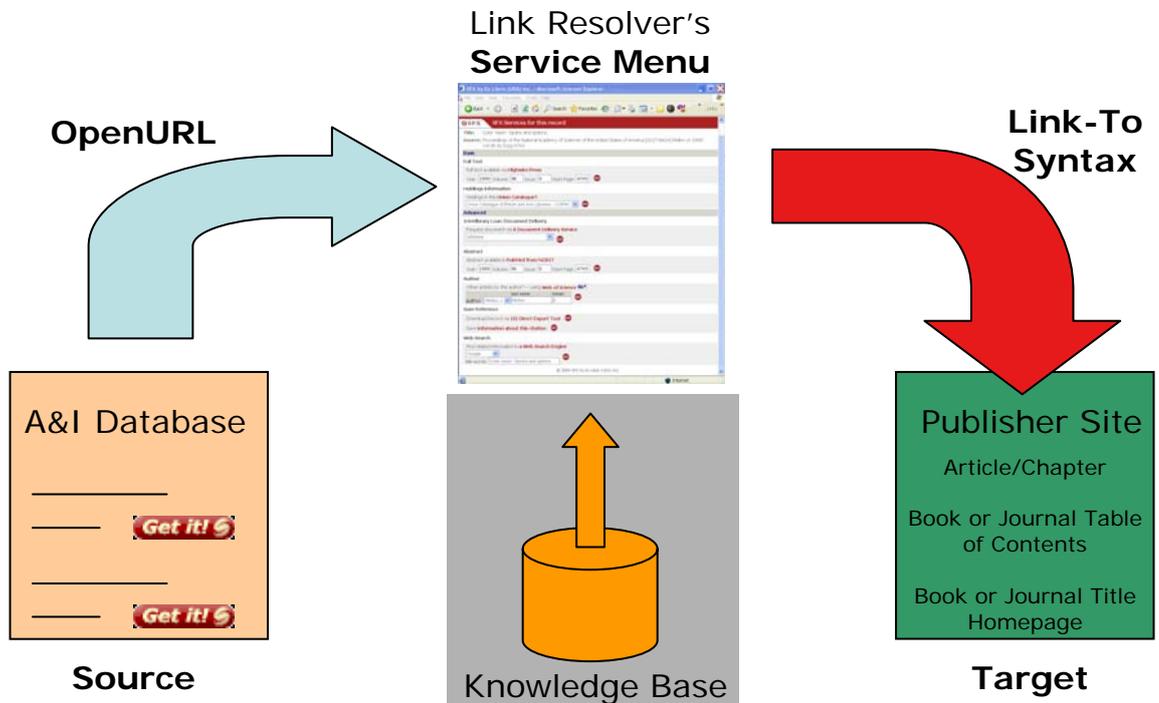


Figure 1: Overview of the OpenURL Linking Framework

The terms referred to in **Figure 1** above are explained in the text below.

### 2.1. THE OPENURL

An **OpenURL**<sup>1</sup> enables the transfer of metadata about an item (a journal article or book, for example) from a resource, where a citation is discovered (for example, an Abstracting & Indexing (A&I) database), to a link resolver. By providing a means to tell another system *what* something is, rather than *where* it is located on the Internet (the function of a normal URL), OpenURLs provide a means for link resolvers to take charge of directing users at particular institutions or organisations to appropriate, subscribed resources for the content, be they in electronic or print form.

This solves a critical problem for librarians: direct URL linking from one publisher's content to another's, including CrossRef DOI-based links<sup>2,3</sup>, has the potential to lead users to resources that are inappropriate for them, i.e. to incarnations of content to which their institution does not subscribe. This is at odds with the librarian's goal of providing access to appropriate (subscribed) content, and has been aptly described as the "appropriate copy" problem<sup>4</sup>. In addition, where multiple subscriptions are held or a number of relevant access points exist, the librarian may desire to nominate the *most* appropriate incarnation of the full text for the user (for

example, should they be directed to the publisher's version or to the licensed incarnation hosted by an aggregator?).

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 appealing outcomes for librarians.

The OpenURL linking syntax was first developed in 2000, an output of research work at the University of Ghent, from which the first commercially available link resolver (Ex Libris' SFX) was to emerge in 2001<sup>5,6</sup>. This linking syntax (known informally as Version 0.1) was, despite its unofficial status, quickly adopted by a significant number of content providers and library systems suppliers. At the same time, the syntax was earmarked for fast tracking to official approval by NISO, the US information standards body. The NISO-approved syntax (informally known as Version 1.0, but officially as Z39.88), was released in 2004. It overcomes some of the limitations of the earlier syntax and is more extensible to other content types. It is therefore intended to replace the earlier syntax. However, the present reality is that both versions of the OpenURL syntax are in use in the scholarly information space today.

## 2.2. SOURCES

Resources that generate OpenURLs for input to link resolvers are known as **Sources**. Most of the key starting points for scholarly research or study – including A&I databases, library web environments and now Google Scholar – are Sources that can interact with the library's resolver in this way. A significant number of primary publishers (and technology service providers to publishers, such as HighWire and Ingenta) have also integrated support for OpenURL links into the reference sections of their online content.

Having located a citation (for example, to a book or journal article) in their resource of choice, the user can query their institution's resolver by clicking on a button or link in the Source application. Often, the button or link presented to the user can be branded or tailored by the institution, to give it a local look and thereby encourage users to click on it. For example, the button for the University of Western Australia (UWA) is branded as follows:



There is a secondary benefit here: a consistent, commonly-presented linking experience for the user, regardless of the resource they choose to commence their research in.

## 2.3. THE KNOWLEDGE BASE AND TARGETS

The OpenURL is, however, only one of two key components that make link resolvers possible. The **Knowledge Base** that underpins the link resolver is also a critical piece of this linking framework as well.

In basic terms, a knowledge base is a directory of where all scholarly content is stored, embracing all of the incarnations of a given title, and how to link to them. For example, a given journal title may be available from a number of different web locations with different coverage ranges: the publisher's own version, including the very latest content, hosted on its own web site; older material hosted by an organisation such as JSTOR; and the licensed database copy, usually with an embargo on the most recent material (available, for example, from EBSCO or ProQuest). Each potential web location is referred to as a **Target**.

Currently, a number of commercial link resolver suppliers independently collect and collate data regarding these different incarnations of online journal and book content from many different information providers, in order to create proprietary knowledge bases for their own products. We refer subsequently to this proprietary knowledge base approach as the "distributed" model. In this model, and in the competitive market place 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 impress the librarian to go with one solution rather than another.

Some libraries have also built knowledge bases themselves as part of 'home-grown' resolver applications (for example, Gold Rush<sup>7</sup>).

Having selected a linking solution from the market place, the librarian (or their resolver supplier) customises 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 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 determines the appropriate options for a particular citation or reference. Does the institution have a subscription to the journal referred to in the OpenURL and does the subscription range encompass the article of interest? Is there more than one point of resolution for the citation? What is the appropriate combination of online full text, print holdings and other services to offer to the user in this specific case? All of these questions can be answered by an interrogation of the knowledge base, dependent, of course, upon its accuracy and comprehensiveness.

Once determined, the relevant options are presented to the user as a menu screen of choices (referred to as the **Service Menu**). Some examples of these menu screens, with the user's options highlighted, are given in **Figures 2 and 3** below.



**You searched for:**

[Refine or alter criteria](#)

**Article:** Progressing towards a biological names register.  
**Author:** Patterson, David  
**Journal:** Nature (London)  
**ISSN:** 0028-0836      **Date:** 17/04/2003  
**Volume:** 422      **Issue:** 6933  
**PMID:** 12700734

**The article is available via the following links:**

|                      |   |   |
|----------------------|---|---|
| Coverage Range       | Links to content                              | Resource  |
| 01/01/1995 - present | <a href="#">Issue</a> <a href="#">Journal</a> | <a href="#">Journals@Ovid Biomedical Collection 3</a> |

**Other options for finding the article:**

Print Materials      [Check the Print Journals A-Z](#)  
                                  [Search the Library Catalogue](#)

Request an Inter-Library Loan   [Click here for details](#)

Figure 2: The link resolver menu for an article from Nature at the University of Wales Institute. This example utilises the Serials Solutions Article Linker software.

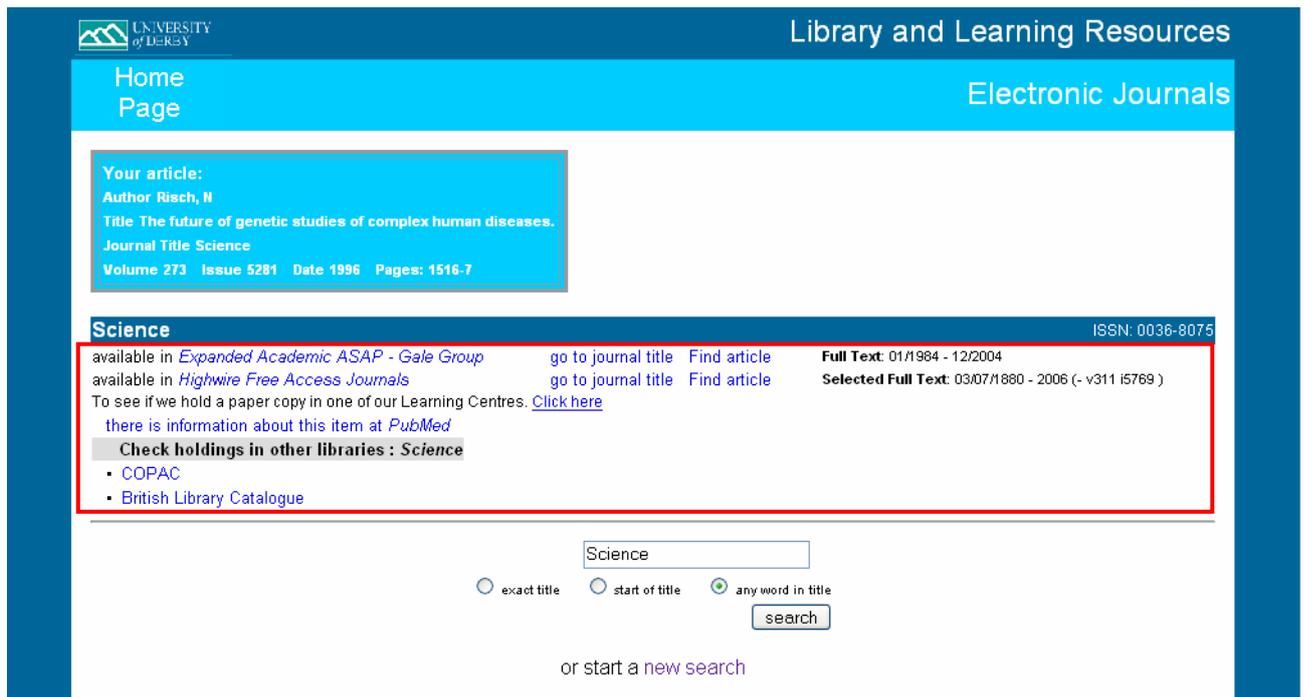


Figure 3: The link resolver menu for an article from Science at the University of Derby. This example utilises OCLC Openly Informatics' 1cate software.

The user selects a destination and the resolver then computes the URL for connecting to the selected Target resource (using the appropriate **Link-To Syntax**). As users may pursue journal article or book chapter citations (in addition to title level citations) the knowledge base must be taught how to

calculate URL links that deliver the user to multiple levels on content sites: title home pages; table of contents pages for books or journal issues; and individual journal article/book chapter pages are the three most common levels.

Both Service Menus and Target links are calculated dynamically, in real-time, by link resolver systems.

## **2.4. UPTAKE OF LINK RESOLVERS**

Through 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. A substantial proportion of academic libraries have already implemented a link resolver, and many others are in the process of doing so.

This rise to prominence could not have been possible without the rapid adoption of the OpenURL linking standard by a significant number of content providers since its appearance in 2000.

## 3. PROJECT CONTEXT

### 3.1. BACKGROUND TO THE STUDY

The significance of the knowledge base to the operation of link resolvers, and by extension to the operation of an increasing number of academic libraries, cannot be understated. As digital collections become more and more critical to libraries, it is essential that the data residing in 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 titles in packages from content aggregators can be inaccurate; and identifying who needs to do what to solve such problems can be difficult<sup>8,9</sup>.

Chen (2004) comments that “just because full-text finding tool vendors update their products regularly does not mean that the lists are actually up-to-date, because full-text finding tool vendors get updates from content providers who have various updating schedules and practices, and thus are of varying quality”. In a similar vein, Wakimoto et al. (2006) report that “the quality of the service could vary widely depending on the accuracy and completeness of the SFX KnowledgeBase. The library has a staff member who “knows better than the vendors do what is in their own databases”, noting that San Marcos reports roughly 30 errors per month back to Ex Libris. Therefore, it took more time to maintain the KnowledgeBase than initially thought”.

For these reasons, the United Kingdom Serials Group (UKSG)<sup>10</sup> invited tenders in July 2006 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 organisations: 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 clarify roles and expectations, and identify performance issues and barriers that need to be overcome to ensure a smooth supply chain of data to the end user.

SIS's bid for the UKSG project was successful and work began in September 2006, with the final project report completed towards the end of January 2007.

### 3.2. OBJECTIVES

SIS's approach was to consult with a sample of stakeholders, drawn from publishers, other content hosts, subscription agents, academic libraries, link resolver suppliers and others, to:

- Understand the present "distributed" supply chain
  - The expectations, roles and relationships between the various agents would be identified and the data flow arrangements mapped and documented
- Firmly establish what is presently not working well
  - Issues, barriers and areas for improvement would be identified
- Consider how the problems identified in the supply chain might be alleviated and who should take action to address them
  - It was agreed at the outset with UKSG that the emphasis in the study would be on practical steps that parties in the supply chain could take to improve on the present environment. This was to embrace best practice recommendations, including quality assurance and timing guidelines. The study would also consider the role of standards in accelerating or automating knowledge base data transfer between the parties.
- Consider an alternative, longer-term approach – a "centralised" knowledge base model revolving around a single repository of content definitions and packages, underpinned by web services and automated publisher data processing
  - The study would seek to establish whether there was enthusiasm for such a model in the community, and – if so – who might be in a position to contribute to such an initiative
- Identify any areas for further research

## 4. METHODOLOGY

### 4.1. STAKEHOLDER INTERVIEWS

The primary research method adopted for the study was one-to-one interviews with stakeholders in the knowledge base supply chain. Many of the interviews were undertaken via telephone, although face-to-face interviews were carried out where possible.

Publisher, content host, link resolver and subscription agent interview targets were identified from SIS's own industry contacts database, giving due consideration to achieving a spread across organisation size and geographic location.

For librarians, SIS drew on its contacts database again but also used web research to establish the link resolver systems in use at a cross-section of libraries and to locate contacts in geographic areas under-represented in its database. SIS sought to identify library contacts that would ensure a spread across library size, resolver system deployed and geographic location.

The initial sample of stakeholders was established in conjunction with UKSG to ensure it was reflective of all players and markets.

Targets were emailed with an introductory message describing the study and inviting their participation via an interview (see **Appendix 1**).

Interviews were typically 45 minutes in length, and were open/unstructured in nature. SIS felt it more appropriate to hear the point of view, concerns and suggestions of the interviewees rather than to steer them through a fixed number of questions (although this was to hand and was used to prompt discussion of topics on occasion).

Prior to undertaking the interviews, SIS prepared – on the basis of its initial understanding – a diagram of the current data flow arrangements between the various stakeholders in the knowledge base supply chain. A discussion of this featured as part of most interviews. Recommendations for adjustments arising from the interviews are reflected in the final version of the diagram and commentary included in this report.

Below is a breakdown by stakeholder type of the interviews completed:

|                          | <b>TOTAL</b> |
|--------------------------|--------------|
| Libraries                | <b>9</b>     |
| Link resolver suppliers  | <b>6</b>     |
| Publishers/content hosts | <b>10</b>    |
| Subscription agents      | <b>2</b>     |
| Others (CrossRef,        | <b>3</b>     |

|                    |           |
|--------------------|-----------|
| EDItEUR, Ringgold) |           |
| <b>TOTAL</b>       | <b>30</b> |

The link resolver suppliers interviewed were:

- EBSCO (LinkSource)
- Ex Libris (SFX)
- Innovative (WebBridge)
- OCLC Openly Informatics (1cate)
- Serials Solutions (Article Linker)
- TDNet (TOUResolver)

Organising interviews with librarians outside of the UK and North America proved challenging. SIS suspects this was due to a lack of recognition of either the SIS or UKSG organisation names amongst the librarians contacted.

Distribution of librarians interviewed:

|               | <b>TOTAL</b> |
|---------------|--------------|
| UK            | <b>5</b>     |
| Europe        | <b>1</b>     |
| North America | <b>3</b>     |
| <b>TOTAL</b>  | <b>9</b>     |

Distribution of resolver systems used by these libraries:

|   | <b>TOTAL</b> |
|---|--------------|
| Ex Libris <i>SFX</i>                    | <b>3</b>     |
| Innovative <i>WebBridge</i>             | <b>2</b>     |
| Ovid <i>LinkSolver</i>                  | <b>1</b>     |
| Serials Solutions <i>Article Linker</i> | <b>1</b>     |
| SIRSI <i>Resolver</i>                   | <b>1</b>     |
| TDNet <i>TOUResolver</i>                | <b>1</b>     |
| <b>TOTAL</b>                            | <b>9</b>     |

## 4.2. LIST SERV MAILINGS

At the end of September 2006, 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, lis-serials). A copy of the message sent is reproduced in **Appendix 2**.

The feedback was disappointing with only four replies received in total, one from a librarian in the USA and three from the UK. This limited feedback has nevertheless been helpful in so far as it has emphasised the same issues gleaned from the stakeholder interviews.

## 4.3. ONLINE SURVEY OF LIBRARIANS

With the emergence of a number of recurring themes from the list feedback and interviews undertaken to the end of November, SIS formulated a short online survey that addressed specific topics (see **Appendix 3**). Invitations to complete the survey were sent by email to library contacts drawn from SIS's own survey database. The aim was to ascertain if the feedback from outside of the UK market concurred with that already captured. No email invitations were therefore sent to UK contacts in the database.

The survey closed in mid-December with 118 respondents. Of the librarians who completed the survey, 90% worked in an academic institution (with or without a research programme).

The geographic breakdown of respondents was as follows:

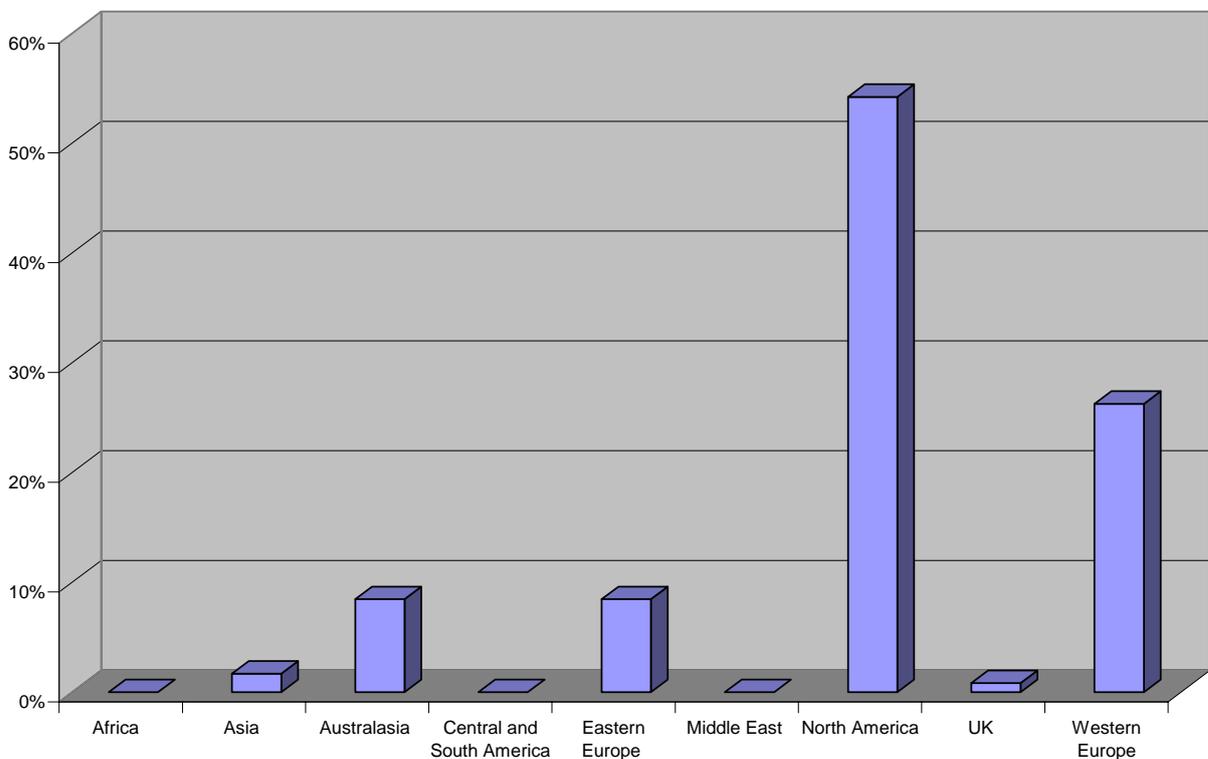


Figure 4: Breakdown of Survey Respondents by Region of World

- 54% of respondents were from North America
- 26% of respondents were from Western Europe (not including UK)
- 8% of respondents were from Australia
- A further 8% of respondents were from Eastern Europe

The breakdown of respondents by link resolver used was as follows:

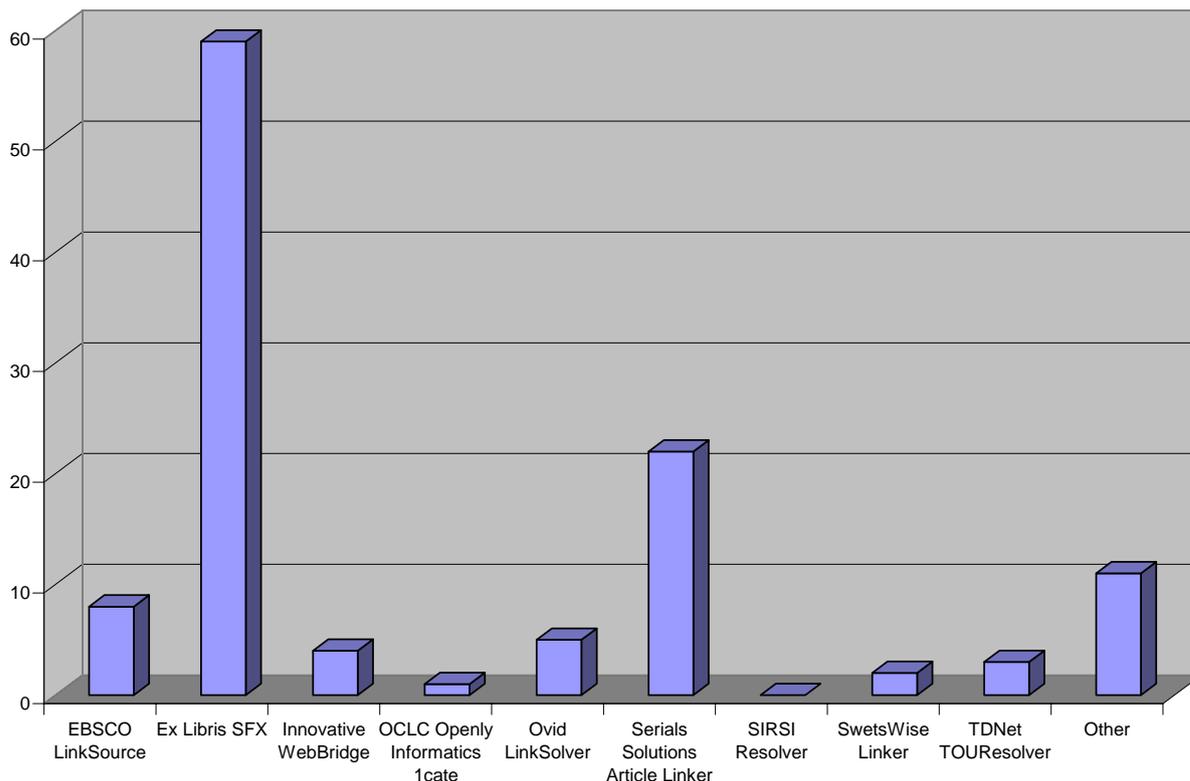


Figure 5: Breakdown of Survey Respondents by Resolver

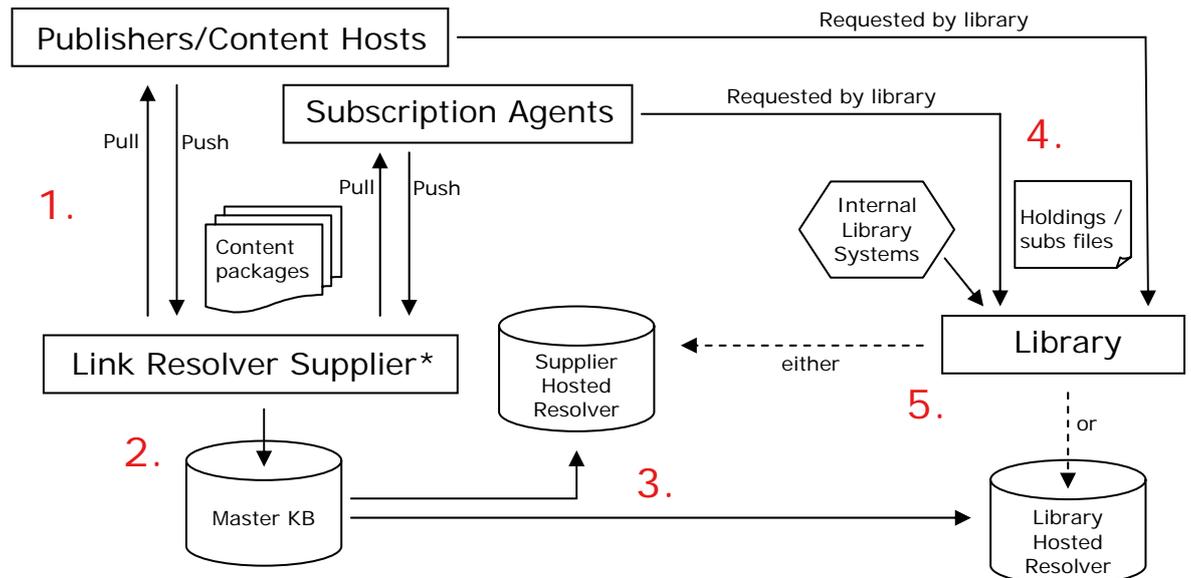
- Ex Libris' *SFX* was the most commonly-used resolver, with 59 respondents (51%) using it
- Serials Solutions' *Article Linker* was the next most popular link resolver with 22 respondents (19%) of the vote
- The "Other" category was the third most popular with 11 votes (10%), including three mentions of Endeavor's *LinkFinderPlus* solution and two mentions of OhioLINK's *OLinks*
- Note: As the online survey did not ask respondents to provide the name of their institution, the data may include respondents from the same institution, utilising the same resolver

In the pages that follow, SIS has referred to the findings from the online survey as they relate to the main narrative of the analysis. SIS has drawn on the data only to emphasise specific points. Graphs or tables of results data are not given (although this information can be provided separately if required).

As a result of receiving an invitation to complete the online survey, a contact at the Colorado Alliance of Research Libraries brought the Gold Rush resolver product to SIS's attention and some feedback from the Gold Rush product manager has been incorporated into the results of the study accordingly.

## 5. DESCRIPTION OF THE SUPPLY CHAIN

### 5.1. GENERALISED SCHEMA FOR KNOWLEDGE BASE DATA FLOW



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

Figure 6: Generalised Description of Knowledge Base Data Flow

#### Supporting Commentary (referring to numbers in chart above):

##### Step 1:

Publishers, content hosts and subscription agents make collection/content package details (and linking syntax descriptions) available to link resolver suppliers. This data typically includes individual titles hosted and their coverage dates. The data is most commonly made available in comma or tab-separated text file format. The data may be pushed to link resolver suppliers or pulled by them. The frequency with which data is updated by content providers and collected by link resolver suppliers varies considerably.

##### Step 2:

Link resolver suppliers process and normalise the collection/content package data and load it into their proprietary master knowledge bases (KBs). A quality assurance layer may or may not feature at this stage. Linking syntaxes for connecting to the content hosting sites are programmed into the supporting software. Additional generic resources may be added directly to the knowledge base by resolver suppliers. For

example, common library catalogues, inter-library loan targets, citation download options etc.

**Step 3:**

Link resolver suppliers distribute the updated master knowledge base (and supporting software) to library implementations. The scale of this task will vary by supplier depending on whether there is support for an ASP solution only (i.e. the solution is hosted for the library by the supplier) and/or support for library-hosted software options. The frequency of update to library implementations is, generally speaking, monthly at minimum, although again there is variation from one supplier to another. In the case of library-hosted resolvers, there may be processing delays due to library/IT staff availability for applying a knowledge base update package.

**Step 4:**

To configure and maintain the resolver, the librarian collects holdings data from internal systems (for example, the library catalogue, the A-Z title list, the Electronic Resource Management tool) and requests/downloads holdings details from subscription agents and individual content providers. Where the content of an acquired package is the same from one library to the next (where it could be said to be 'standard'), the collection of title level details is not usually necessary. For example, aggregator database products and common subject back file collections.

**Step 5:**

'Standard' packages are simply activated in the knowledge base (the library usually trusts the resolver supplier to have described the package correctly). Holdings data that reflects local library conditions is uploaded to the resolver (either by the library or its resolver supplier) to customise other content packages and resources. This data will need to be processed and normalised to the format expected by the resolver system prior to upload.

**5.2. ROLES AND RELATIONSHIPS MATRIX**

The following matrix is an *idealised representation* of the current roles and relationships between stakeholders in the knowledge base data flow, drawn from interview discussions:

| Stakeholder              | Knowledge Base Supply Chain Role(s)                                      | Relationship With      | Stakeholder's Expectation of Relationship  |
|--------------------------|--|------------------------|--|
| Publisher / Content Host | To make collection data description available to Link Resolver Supplier. | Link Resolver Supplier | That Link Resolver Supplier will process/upload data in a timely fashion and distribute on to the Library. |
|                          | Where relevant, to make subscription files/holdings details              | Library                | That library will request/download data as and when it needs   |

|                    | <p>available to Library for use in customising the Publisher/Content Host's knowledge base target to reflect local conditions.</p> <p>To provide details of a suitable inbound linking syntax for delivering Library users to the content (Target).</p> <p>Where relevant, to provide details of OpenURL support (Source).</p>  | <p>Link Resolver Supplier</p> <p>Link Resolver Supplier and Library</p>  | <p>it.</p> <p>That Link Resolver Supplier will configure its system to utilise the linking syntax for users connecting to the Publisher/Content Host's material.</p> <p>That Link Resolver Supplier will check its system integrates satisfactorily and that Library will enable OpenURL linking on the Publisher/Content Host's web site for its users.</p>   |
|--------------------|---|--|--|
| <b>Stakeholder</b> | <b>Knowledge Base Supply Chain Role(s)</b>  | <b>Relationship With</b>   | <b>Stakeholder's Expectation of Relationship</b>   |
| Subscription Agent | <p>To make collection data description available to Link Resolver Supplier [for agent gateway product].</p> <p>To make subscription files/holdings details available to Library customer for use in customising knowledge base targets to reflect local conditions.</p> <p>To provide details of a suitable inbound linking syntax for delivering Library users to the content (Target).</p> <p>To provide details of OpenURL support (Source).</p> | <p>Link Resolver Supplier</p> <p>Library</p> <p>Link Resolver Supplier</p> <p>Link Resolver Supplier and Library</p> | <p>That Link Resolver Supplier will process/upload data in a timely fashion and distribute on to the Library.</p> <p>That library will request/download data as and when it needs it.</p> <p>That Link Resolver Supplier will configure its system to utilise the linking syntax for users connecting to Subscription Agent's content.</p> <p>That Link Resolver Supplier will check its system integrates satisfactorily and that Library will enable OpenURL linking on the Subscription Agent's web site for its users.</p> |
| <b>Stakeholder</b> | <b>Knowledge Base Supply Chain Role(s)</b>  | <b>Relationship With</b>   | <b>Stakeholder's Expectation of Relationship</b>   |
| Link Resolver      | Gathering,  | Publisher /  | Data provided is:  |

| Supplier    | <p>normalisation and processing of collection data for loading into master knowledge base.</p> <p>Gathering and configuration of inbound linking syntaxes to accompany collection data for Targets.</p> <p>Checking (and possibly documentation) of OpenURL support in Sources.</p> <p>Distribution of updated knowledge base data to Library.</p>  | <p>Content Host and Subscription Agent</p> <p>Publisher / Content Host and Subscription Agent</p> <p>Publisher / Content Host, Subscription Agent and Library</p> <p>Library</p> | <p>accurate/current; consistently laid out from one update to the next; sufficiently descriptive (title identifiers, coverage information); structured (comma or tab-separated file format is typical).</p> <p>Mechanism for data to be automatically sent or collected is in place.</p> <p>That linking syntaxes are: flexible (support linking to different levels); well described; do not rely on internal identifiers; do not change over time.</p> <p>Link Resolver Supplier may provide documentation regarding OpenURL support by content sites to the Library as part of its service.</p> <p>That Library will process the update in its own time (if such a step is necessary) and take any other manual action resulting from changes in the knowledge base contents (for example, configure new targets, make live new titles added to an existing subscribed target).</p> |
|-------------|---|--|--|
| Stakeholder | Knowledge Base Supply Chain Role(s)   | Relationship With  | Stakeholder's Expectation of Relationship  |
| Library     | <p>To activate 'standard' subscription targets in the knowledge base (for example, aggregator database products).</p> <p>To collect subscription details from internal systems (library catalogue, A-Z list, ERM tool etc) for customising other knowledge base targets to reflect local conditions.</p> <p>To collect subscription details from Publisher/Content Host for customising</p> | <p>Link Resolver Supplier</p> <p>Other library systems [Internal]</p> <p>Publisher / Content Host</p>  | <p>Library expects complete accuracy in the data for these targets. Does not have time to check.</p> <p>Data should be: accurate; current; structured; easy to collect etc. These are all issues internal to the library and the suppliers of its own software systems.</p> <p>Data is: accurate/current; consistently laid out from one update to the next; sufficiently</p>  |

|             | <p>the relevant knowledge base target to reflect local conditions.</p> <p>To collect subscription details from Subscription Agent for customising some knowledge base targets to reflect local conditions.</p> <p>To receive and apply (or have applied) knowledge base updates from Link Resolver Supplier.</p> <p>To process knowledge base updates from Link Resolver Supplier in a timely fashion for Users [if such a process is necessary, only likely if locally hosting the solution].</p> | <p>Subscription Agent</p> <p>Link Resolver Supplier</p> <p>User</p> | <p>descriptive (title identifiers, coverage information); structured (comma or tab-separated file format is typical).</p> <p>Mechanism for online collection is in place.</p> <p>Data is: accurate/current; consistently laid out from one update to the next; sufficiently descriptive (title identifiers, coverage information); structured (comma or tab-separated file format is typical).</p> <p>Mechanism for online collection is in place.</p> <p>Data is accurate and current. Library expects Link Resolver Supplier to have checked the data it received from the Content Hosts before distributing it further. Updates should be provided to the Library on at least a monthly basis.</p> <p>The provision of an accurate and reliable service to the end customer.</p> |
|-------------|--|---|---|
| Stakeholder | Knowledge Base Supply Chain Role(s)  | Relationship With   | Stakeholder's Expectation of Relationship   |
| User        | To use the OpenURL-based linking service provided by the library.  | Library   | User expects accuracy in service menus and links at all times.  |

### 5.3. ADDITIONAL REMARKS

It is important to stress that both the data flow schema and roles/relationships matrix presented above are a generalised representation of the current situation. There are a number of issues and problems in the present environment, and these are discussed in **Section 6**.

Here, it is worth touching on a number of potential complexities to the data flow and to the relationships between parties that are not detailed in the documentation above:

- Some publisher products are not sold through subscription agents, meaning that agents are not in a position to provide data to libraries to assist knowledge base localisation.
- Link resolver vendors may have more complex relationships in place for sourcing knowledge base data than those described. Some vendors completely outsource the collection of knowledge base data to a third party, and are therefore more detached from the collection and data normalisation activities than others. In addition, some suppliers of link resolver tools license the entire product from a third party – both the knowledge base data and software functionality are effectively outsourced to a specialist.
- On the library side, there is scope for further complexity as a result of the interplay of various library systems. This is only casually alluded to above. There is the potential for a number of data staging posts prior to upload to the link resolver knowledge base. For example, some libraries populate their resolver's knowledge base entirely from the A-Z list product of another library systems supplier, sourcing data for this from yet other applications. With the release of Electronic Resource Management (ERM) products into the library market in the last 1-2 years, there is an increasing level of interaction (perhaps as frequently as daily) between the knowledge bases for different library products: from the ERM tool to the link resolver and A-Z list and vice versa, for example.
- It should also be noted that interplay between a consortia-level knowledge base and those of individual member libraries is perfectly feasible in the model. Some link resolver suppliers have developed complex mechanisms and arrangements to enable library consortia to operate effectively, with link resolvers querying one another to aggregate the list of appropriate services for a given OpenURL.
- Who has responsibility for localising the knowledge base can also be different from one set-up to another. Some resolver suppliers provide a complete toolkit to libraries in addition to a global knowledge base, effectively enabling the library to localise everything from scratch itself. Other arrangements see the library passing holdings details onto their resolver supplier, with the latter uploading the information to the knowledge base on the library's behalf. It should be noted that, whatever the arrangement, unless the resolver system in use is that of the subscription agent for a given library, the agent tends to be somewhat removed from the knowledge base localisation process (we return to this subject later).
- Libraries may or may not have the means (in terms of system used or in terms of staff resource) to create knowledge base entries afresh themselves, i.e. creating a completely new target in the system as opposed to simply localising a pre-existing one.
- Furthermore, libraries can play an important role in sending details and data files of a content provider's collection to their link resolver supplier, to ensure targets are added to the global knowledge base system. This is

despite the fact that the link resolver supplier cannot (for whatever reason) source collection data directly from the provider itself.

#### 5.3.1. CrossRef

Interaction with CrossRef is not shown in the schema but some comment on it is warranted.

Whilst some integration between link resolvers and CrossRef exists already, it is limited to 'on the fly' querying of the CrossRef system to establish the DOI for a given OpenURL item or to retrieve metadata about an item (for example, ISSN, volume, issue, start page) based on the DOI sent in the OpenURL. This integration is useful in ensuring the fullest possible set of linking services is displayed to the user of the resolver.

There is also the means, should a library wish to utilise it, for DOI links in online content to connect to an institutional resolver, i.e. for the DOI link to be transformed into an OpenURL, and for the DOI system to become a Source.

#### 5.3.2. Google Scholar

As discussed earlier, a resource is deemed a Source if it can (a) generate OpenURLs for citations or references in its content, and (b) send these OpenURLs to the relevant link resolver for a given user. What happens next, in terms of the menu of services that is presented to the user, is out of the Source's control. To this extent, the Source is linking blind to the link resolver system – it has no means of knowing what the menu will offer the user.

Google Scholar (and a small number of other Sources) has been unwilling to accept this arrangement, wishing to control the user experience in its own interface a little more finely. This has been achieved by requiring libraries to relay details of the full text titles and subscriptions activated in their resolver knowledge bases to Google (in an XML file format), such that Google can query this information during search results page generation and highlight those results for which the Scholar user has full text rights. The user is then connected through to the institution's link resolver in the normal way from the search results page (i.e. via an OpenURL), but with the added benefit of prior knowledge of the outcome.

This is an additional data flow that is not represented in the schema above, but is shown separately below (See **Figure 7**).

A link resolver supplier may provide a mechanism to facilitate the data transfer to Google Scholar, with the library responsible for operating it, or the resolver supplier may carry out the data transfer work itself as a service to the library customer.

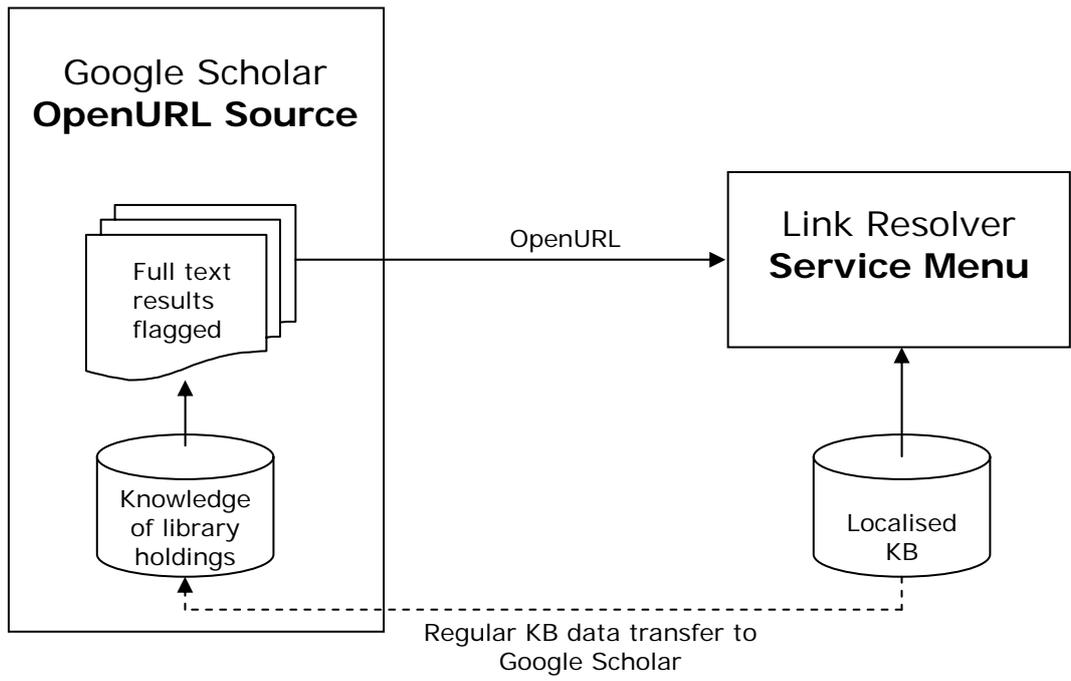


Figure 7: Knowledge Base Data Interaction with Google Scholar

## 6. ISSUES AND BARRIERS

### 6.1. LACK OF AWARENESS

Whilst some information 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 realising that this is a desirable thing to do.

All of those interviewed agree that awareness has certainly improved in recent years, but the online survey results indicate that almost a third of librarian respondents felt that between 11-20% of content providers were still unaware of link resolvers and do not understand what they should do to work with these systems. A further 39% considered 1-10% of the market to be similarly unaware. SIS thinks it likely that most survey respondents answered this question with reference to the providers they manage to speak to/meet with (the top 100 or 200, say). From interview discussions, our experience suggests that there is a large body of small publishers who know very little or nothing at all about link resolvers. The numbers reported in the survey probably under-represent the true picture as a result.

There is, therefore, a very definite need for further education in the market place, specifically with smaller publishers and content hosts in certain subject disciplines (for example, and especially in, law). However, even the larger publishers themselves admitted during interview that they have only a limited understanding of resolvers and lack an overall view of this technology space. One large publisher told SIS that it works closely with CrossRef, and - believing this to be sufficient - lets the link resolvers do the rest. Another medium-sized publisher told us that only a few select staff at its organisation understood link resolvers and appreciated the significance of making arrangements with suppliers. Several publishers commented that data flow arrangements of this kind tend to fall between departments and that – as a result – there can be a lack of clear ownership and drive.

Content providers need to hear how the collection data they provide to link resolver suppliers impacts the library user and impacts them. Amongst some providers there may be a tendency to view data supply to link resolver suppliers as only helping the latter to make a sale. Correcting this view and encouraging publishers to view link resolver suppliers as a key distribution partner (a customer even) that can critically affect the visibility of their content, and therefore its impact amongst readers and authors, would be very helpful in improving the current situation.

### 6.2. LACK OF CO-OPERATION

Some content providers have stated to link resolver suppliers that they are unwilling to provide data feeds. This is perhaps because they view their

collection data as proprietary or commercially sensitive. They need educating as to why they should not feel that this is the case and what the benefits of working with resolver suppliers are.

To ensure its resource continues to be used as a searchable database, rather than having users 'jump' in and out of it at a granular (title/article) level, a content provider might perhaps avoid or limit integration with link resolver suppliers. A significant number of interviewees suggested that this might be the case with some providers. If so, there is again a misconceived threat here that needs tackling through education.

Publishers are notably more open and co-operative than they were five years ago, but the level of co-operation *in some cases* is still not all that it might be. Resolver suppliers have sometimes found themselves reluctantly forced into manually copying information about a publisher's collection from the publisher's web site, due to a lack of any structured list being made available. Alternatively, they may have to rely entirely on libraries to provide the collection data for certain providers in order to add it to their knowledge bases.

However, and significantly, many publishers also 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.

There are some more obvious reasons for a lack of close communication where a subscription agent's data about a library's holdings and a competing party's link resolver solution are concerned. There are some significant competitive issues that obstruct data sharing between stakeholders in the supply chain and they will not easily disappear.

### **6.3. INACCURATE AND INCOMPLETE DATA**

For the many content providers that do supply collection data to link resolvers and holdings data to libraries, the accuracy, completeness and ease of obtaining this information is critical. At the end of the day, libraries are depending on the data provided to offer a reliable service to their patrons.

The completeness and accuracy of collection information being transferred to link resolver systems varies considerably from one content host to another. In connection with a lack of awareness, some content hosts may not be aware of the precise use to which their data feeds are put by link resolver suppliers and libraries, and hence of the need for accuracy and currency. The types of errors in these data feeds have been well reported elsewhere<sup>8</sup> but include missing titles, titles listed in error, wrong identifiers for titles (ISSN, ISBN), incorrect coverage information and incorrect embargo information.

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 fix the problems at source. Many have concluded that full text aggregators in particular focus their energies in

other areas and metadata accuracy is never (voluntarily at least) going to be of high concern to them. As a result, substantial rule tables have been separately developed by resolver suppliers to fix updates from content providers that are guaranteed to contain errors seen and addressed before. Other content providers are happy to fix errors in their data but operate in a rather reactive manner, waiting for the resolver suppliers or libraries to inform them of problems rather than identifying the problems prior to initial supply.

To put the scale of the problems here into perspective, 7% of the librarians in our online survey reported that their perception was that over 25% of the data in their resolver's knowledge base was inaccurate. The largest proportion of respondents (40%) felt the errors to be between 6-10%, with 29% voting for 11-25%. The problem of data inaccuracy, whilst impacting only a minority share of the content within knowledge bases, is nevertheless still a significant one for librarians. A small number of errors and problems might be enough to lead a user to abandon the library's linking solution in favour of something else.

Chen (2004) remarks that "Content providers need to realize the serious consequences of misinformation in serials management systems, OpenURL link resolvers, and imported e-journal MARC records. Libraries should use the quality of full-text source lists as an important selection criterion when they shop around for full-text databases". This course of action, whilst appealing to the librarian, is not however always practical. For example, where a resource providing a poor source list is a critical one as far as a faculty is concerned. A difficult balancing act may be the result for the librarian.

#### **6.4. CONTENT PACKAGE ISSUES**

The complexity of some publisher package deals and the scope for variation in packages purchased by libraries (for example, as a result of specific national deals, such as NESLi2) can create problems and overhead for both link resolver suppliers and librarians alike.

There is patchy support for tracking national, bundled deals (with subtle variations in content) in resolver knowledge bases. This may be because data is not forthcoming from publishers. One librarian interviewed felt very strongly that there is a huge responsibility on those negotiating national deals to make demands of publishers in this area - to require them as part of the negotiation process to make a structured list of the contents of a deal available to libraries, subscription agents and link resolver suppliers.

That said, one publisher commented to us that they have such a proliferation of products and packages that it is simply not realistic to expect structured lists to be made available for all of them. Similarly, one link resolver supplier reported on the difficulties a large publisher is known to have in outputting title data in package groupings for them.

In the library survey, 62% of respondents either agreed or strongly agreed that some publishers have difficulties in making title information for

consortia deals available to libraries and their link resolver systems. Only 10% disagreed or strongly disagreed.

And yet, the package issue is also a somewhat awkward one for link resolver suppliers. They would like to oblige their library customers, but need the data to come in a structured form from somewhere. Tracking every national deal from every publisher listed in a knowledge base also sounds a very unenviable task if the data cannot be trusted to be correct.

A related theme is the current lack of common identifiers for content packages. A global registry for assigning identifiers to content packages, identifiers that would persist between different link resolver solutions and across different library systems (from link resolver to ERM system, for example), could bring significant benefits. Ted Koppel, Verde Product Manager at Ex Libris, recently cited the need for unique collection identifiers (“like an ISBN per e-package”) as number two in his wish list of top five most-needed information standards<sup>11</sup>.

A recurring outcome here then is manual work for the library (and/or its resolver supplier) in configuring knowledge base targets that accurately reflect the deal struck with a publisher. This is clearly undesirable, and – given that publishers are not going to stop being creative when it comes to cutting deals with different library consortia – something should be done to ease the burden of supporting them in knowledge bases.

## **6.5. JOURNAL TITLE CHANGES AND TRANSFERS BETWEEN PUBLISHERS**

Journal title changes and transfers between publishers can be a significant area of difficulty for link resolver suppliers and libraries.

When given the statement “Publishers are doing enough in managing journal title changes and journal transfers”, 58% of respondents to the survey either disagreed or strongly disagreed. Nobody strongly agreed with this statement and only 19% agreed. One librarian told us that “each publisher seems to have its own way of dealing with title changes: for example, cross-referencing the new title and the old title while keeping separate web sites for the content (good); consolidating all content under the new title but keeping cross-references from the old title’s web site (worse); consolidating all content under the new title and completely removing all reference to the existence of a former title (worst)”.

Resolver suppliers comment that such title changes are described inconsistently in data files from content providers, and that this area alone constitutes a significant part of the quality control work they undertake.

It is therefore felt that a transparent, consistent, and logical approach to this topic is badly needed.

The TRANSFER initiative<sup>12</sup>, looking to formulate a code of practice for journal transfers between publishers, is a welcome development by libraries and link resolver suppliers.

## 6.6. RESPONSIBILITY FOR DATA QUALITY

Libraries rely on the data within the knowledge base that their resolver supplier provides to be accurate. Whilst link resolver suppliers may make improvements to the data they receive from content providers (discussed above) they are fundamentally in the hands of the content providers to supply good data to them in the first place. When one considers the scale of the knowledge base task, it is not surprising that the contents of every data file processed may not be checked line by line.

However, this should clearly not mean that librarians have to take responsibility for data quality. Firstly, it is clearly inefficient for a large number of librarians to identify and report the same problems to their resolver supplier. Secondly, responsibility for the errors identified by librarians may not be claimed quickly, leading to frustrating delays in addressing the problem: is the link resolver supplier or the publisher responsible?

Again, there is a greater need for understanding, communication and co-operation between content providers and link resolver suppliers to solve these quality assurance problems at source.

## 6.7. LACK OF DATA STANDARDS

A number of the problems in the supply chain stem from the fact that there is a lack of an agreed standard format for transferring collection/holdings data between stakeholders.

Most content providers and subscription agents that supply data in a structured format to link resolver suppliers and libraries today do so using comma or tab-separated files (with a few supplying data in XML). But there is no consistency in format and field labelling from one provider to another, creating manual effort for the library, and a huge data normalisation problem that is duplicated across each of the link resolver suppliers. Given the scale of the normalisation task, it is again not hard to understand how examining the actual content of the data files processed may take second place to simply getting the data uploaded to the knowledge base to begin with.

Libraries customising the 'out of the box' knowledge base themselves may face a significant overhead in terms of reformatting holdings files downloaded from publisher/subscription agent sites to a format that can be loaded into their knowledge base. Issues faced along the way may be insufficient or inappropriately formatted date and enumeration details as well as missing or wrong identifiers (ISSN, ISBN).

## 6.8. TIMING ISSUES

How often content providers update their collection data, how often is it pushed to or collected by resolver suppliers and how quickly is it distributed onto libraries all varies considerably in the supply chain today.

Some publishers may not need to offer frequent updates if their content is static for long periods of time, but others should refresh the data regularly and consistently. There is no common benchmark however and our research suggests that monthly updating is probably the typical norm for content providers.

A monthly collection and processing cycle (of the data from content providers) is also the norm for link resolver suppliers.

Distribution onto libraries varies, partly depending on model of operation (supplier-hosted ASP solution, or locally-hosted library implementation). Some link resolver suppliers achieve real-time or overnight updating.

Many of the librarians that utilise a service with a monthly knowledge base update cycle are really seeking a much more rapid passage of data through the supply chain – from publisher update to local knowledge base update – to retain currency and accuracy for users.

The lack of adherence to a standard data format is, again, one significant obstacle to more frequent knowledge base updates.

## 6.9. INBOUND LINKING ISSUES

There is a lack of a common linking standard for routing users to content on publisher sites. 71% of respondents to the online survey either agreed or strongly agreed that the effectiveness of link resolvers is hampered by the lack of a standard linking syntax for connecting to publisher web sites.

Not all target resources support the ideal: journal article or book chapter level linking. Some only support it via DOI, which is not really sufficient for all linking scenarios.

Again, work is created, as a result, for all of the link resolver suppliers - to configure their knowledge bases to cope with an ever-growing number of different linking syntaxes and methodologies for delivering users to target sites.

Librarians may also need to work to share information to try to get around an apparent restriction on a publisher's site – to find a way to link to the target at a deeper (more appropriate in their eyes) level. The alternative may be to live with a less than satisfactory linking experience for their users.

Whilst adherence to a common standard may be unachievable, adherence to a common approach, combined with greater publicity of available syntaxes by publishers, would be very helpful.

From the content provider's perspective, there is no common statistical measure for determining which link resolver systems are creating inbound links accurately and which are not. If a publisher has integrated with a link resolver solution, it may wish to be assured that the URLs coming from that source are computed to a high degree of accuracy, to avoid the user being taken to an error page and believing it to be the publisher's fault. This study has established that some content providers have requested

identifiers in inbound links to indicate the originating resolver system and this approach could perhaps be more broadly adopted in the future.

## 6.10. OPENURL ISSUES

OpenURL issues are somewhat tangential to the focus of this study, but – as the critical input to everything that follows in the user experience with link resolvers – we have considered the current problems in this area also.

72% of respondents to the online survey either agreed or strongly agreed that a significant problem for link resolvers is the generation of incomplete or inaccurate OpenURLs by databases (for example, A&I products). OpenURLs may be broken on account of insufficient or incorrect metadata that leads to erroneous results in the link resolver's service menu or prevents the resolver from creating a sufficiently deep link to a target site. One librarian interviewed commented that his experience with some sources was so bad that he refused to enable OpenURL links from them as he did not wish to expose his end users to the problems.

There is a potential for disconnect between publishers and A&I companies that may have a bearing on OpenURL problems. Items such as combined issues and supplements may not be tracked identically and journal title changes may also be managed differently. Such discrepancies may explain some of the problems in accurate OpenURL generation and onward linking from resolver menus.

Some referring sources need prompting to display institutional links prominently and others do not permit the use of institutional logos (for example, Google Scholar), which library users may expect to see.

OpenURL compliance has not penetrated to all potential referring sources, particularly the smaller ones. Some vendors that state they are OpenURL compliant are not overly clear about their support for it, and do not provide a straightforward mechanism for enabling OpenURL links to the library's link resolver (via an easy-to-use administration interface, for example).

Indeed, there remains a misunderstanding in some quarters of the market as to the difference between the (outbound) OpenURL syntax and inbound linking syntaxes. Some content providers have adopted an OpenURL-based syntax for inbound links. Whilst technically sound, this was never the original intention or purpose of the OpenURL and its use in this way is a cause of some confusion.

During interview, some content providers admitted their lack of understanding of components of the OpenURL syntax, and their confusion as a result of the existence of two OpenURL linking standards. A number remarked that the Z39.88 documentation is quite complex and hard to understand. The earlier syntax is appealing as a result of its relative simplicity. One publisher that has adopted OpenURL 0.1 saw no compelling reason to move to Z39.88. We suspect other publishers may feel the same.

## 6.11. THE ROLE OF THE SUBSCRIPTION AGENT

Many librarians commented that the localisation of resolver knowledge bases to reflect individual library conditions was a very time-consuming area of work for them. There is, as a result, a desire from many to see subscription agents playing a far greater (and more direct) role in the localisation task. In the librarian survey, 74% of respondents either agreed or strongly agreed that subscription agents could assist libraries further by working more closely with link resolver suppliers to configure knowledge bases. Only 6% of our sample disagreed or strongly disagreed with this statement.

One librarian thought it perverse that he was having to source holdings data direct from content providers for loading into his knowledge base when his subscription agent should be able to do this or have this level of detail already in its systems. Another wondered if the subscription agent could be the "missing link in the chain" and provide a "value-added service of aggregating holdings information from content providers, standardising its format and routing it onto the library's link resolver supplier". Another commented that agents should be focusing their efforts on providing mechanisms for libraries and others to collect holdings data from them.

The idea of becoming more involved in knowledge base localisation via automated (rather than manual) solutions was a welcome one to the subscription agents interviewed. However, it was felt that agreements with libraries and third parties would have to be very carefully structured to protect the data the agent would make available in this way (i.e. information about an individual library's subscriptions as managed by the agent). The issue of commercially sensitive data finding its way to a direct competitor is an awkward one. And yet smaller libraries with limited staff resources would in particular seem to be looking for some kind of service from the agents in this area.

A further potential complication for an agent-led service would be library deals struck with publishers without agent knowledge or involvement. For example, back file packages purchased directly for a one time fee. The degree of assistance offered would obviously be limited to the agent's knowledge of the library's subscriptions and holdings only.

## 6.12. BROADENING OF KNOWLEDGE BASES

A number of interviewees commented that the scope of present-day resolver knowledge bases was largely limited to full-text journal and (a growing amount of) book content, and that this will surely need to change.

The e-resource arena, which started in serials, is now quickly moving beyond them. Image databases, A&I database contents, encyclopaedic entries and Open Access articles published in hybrid journals were all mentioned as content types that need to be addressed in the future by stakeholders in the supply chain.

### **6.13. SUMMARY: OPENURL COMPLIANCE IS NOT KNOWLEDGE BASE COMPLIANCE**

It is important to state that many academic librarians are happy with their link resolver solution and reasonably confident in the knowledge base that underpins it. It cannot be denied that link resolvers have already proved themselves to be a great step forwards for electronic resource management in the six years or so since they first appeared on the library scene.

And yet, from the list of issues and problems described above there is a general sense of an ad hoc feel to the current arrangements in the knowledge base supply chain. The data flow works by and large (certainly for most of the large content providers), but seems to have developed in an unstructured way, creating confusion for some and challenges/issues for others. As well as solving problems, resolver technologies have created new problems and manual effort for academic librarians and their suppliers.

It could be said that whilst the community's attention has been mostly focused on what it means to be *OpenURL compliant*, a code of practice and information standards to ensure *knowledge base compliance* and the efficient transfer of data through the supply chain have been sorely absent and overlooked. We turn to this subject in the next section of the report.

## 7. RECOMMENDATIONS

### 7.1. "COUNTER" FOR KNOWLEDGE BASE AND OPENURL COMPLIANCE

A number of the issues and barriers in the present knowledge base supply chain stem from a lack of awareness or understanding of link resolvers and OpenURL linking. Coupled with this is the lack of a consistent and transparent set of guidelines for parties in the supply chain to adhere to.

SIS believes these issues could be most effectively addressed through education and communication via a mirror organisation to that which operates in the usage statistics space (Project COUNTER<sup>13</sup>). This organisation, which the UKSG could take the lead in establishing, would seek to bring stakeholders together to define a code of practice for effective participation in the knowledge base supply chain. Libraries could then point content providers 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 is appropriate, worthwhile and valued. Greater transparency in this area, they argue, would increase uptake and involvement from publishers and other content hosts.

Project COUNTER is a good role model to adopt, chiefly as a result of its open, inclusive and interactive nature: publisher, library and intermediary representatives were all involved from its very inception. In addition, the deliberate limiting of the initial scope of the initiative (to a code of practice for journals and databases), which has more recently been systematically widened as awareness has grown, has been another factor in its success.

#### 7.1.1. Code of Practice

SIS believes therefore that an initial scope should be limited to the promotion of a straightforward code of practice to the various stakeholders in the knowledge base supply chain. We provide the parameters and probable values for four major areas of recommendation (A-D) within a code of practice below, as a starting point for further discussion:

#### **A. Knowledge Base Compliance for Publishers and Content Hosts**

- *Collection data supply to link resolver suppliers:*
  - File format – must be a comma- or tab-separated values file
  - Agreement on file layout – mandatory fields (ISSN, eISSN, ISBN, book/journal title, coverage details etc) and sequencing of fields

- Agreement on content of fields – common approach to dates, enumeration information and embargoes, for example
- Scope – all titles hosted by content provider to be listed in one file (mandatory); relevant sub-collections/content packages in separate files (desirable – see below)
- Frequency of data update – monthly is mandatory, weekly is desirable (will not need to apply to those whose content changes infrequently)
- Collection mechanism – must be web or FTP-based
- *Holdings data supply to library customers:*
  - File format – must be a comma- or tab-separated values file
  - Agreement on file layout – mandatory fields (ISSN, eISSN, ISBN, book/journal title, coverage details etc) and sequencing of fields
  - Agreement on content of fields – common approach to dates, enumeration information and embargoes, for example
  - Sub-collection/package information (see below)
  - Frequency of data update – overnight is mandatory, real-time is desirable
  - Collection mechanism – must be web-based, from within a library administrator area or similar
- *Inbound linking syntax:*
  - Note: As many publishers have developed deep inbound linking syntaxes that link resolvers are already supporting, the guidelines below should only need to apply to those developing a syntax for the first time, or to those whose existing syntax is in some way deficient.
  - A flexible (multi-level) linking syntax that relies on standard, common metadata elements is desirable. Common metadata elements would include ISSN, ISBN, year, volume, issue and start page number, for example. The option to quote author names is useful when there is the possibility of journal articles starting on the same page in an issue. Poor metadata elements are internal numeric or alphabetic identifiers for titles/articles/chapters that cannot be derived (calculated) from other information in any way, and should be avoided.
  - The linking syntax is best expressed as an arrangement of key-value pairs. Keys can be thought of as headings or labels (for example, 'year') against which values are listed (for example, '2001'). Thus, a key-value pair is the combination of the two elements ('year=2001'). Ampersands are used as separators between key-value pairs in a URL string. For example, <http://www.publishing.com/link?issn=0123-4567&year=2005&volume=10&issue=1&startpage=59>.
  - Support for DOI-based inbound linking is welcome but insufficient on its own
  - A rating system could be employed to enable librarians to make comparisons between publishers' linking syntaxes (like that at: [www.exlibrisgroup.com/sfx\\_targets.htm](http://www.exlibrisgroup.com/sfx_targets.htm))
  - Documentation requirements (promotion of the syntax is critical)
  - Agreement to provide three months' written notice of any planned change in inbound linking syntax

## **B. Knowledge Base Compliance for Subscription Agents**

- *Collection data supply to link resolver suppliers:*
  - As above, except sub-collections/content packages in separate files can probably be ignored (the agent will describe its complete collection only)
- *Holdings data supply to library customers:*
  - As above, except replace or expand sub-collection/package information with publisher/content host details
- *Inbound linking syntax:*
  - As above

## **C. Knowledge Base Compliance for Link Resolver Suppliers**

- Timely upload of collection data from publishers/content hosts/subscription agents – mandatory frequency is monthly, a higher frequency is desirable
- Commitment to subscribe to all relevant email alert services (and any other sources of relevant information) to ensure knowledge of new databases/new packages/journal title changes/journal publisher changes in advance and independently of the library community
- Timely distribution of knowledge base data onto library customers – mandatory frequency is monthly, a higher frequency is desirable
- Timely processing of localisation data from libraries (where relevant to supplier's business model) – overnight is mandatory, real-time is desirable

## **D. OpenURL Compliance for Content Providers**

- Define the version of the OpenURL syntax to adopt and the mandatory fields to populate – greater publicity of the Z39.88 "Implementation Guidelines"<sup>14</sup> may be sufficient, or the preparation of a shorter, simpler guide derived from this document may be beneficial
- Minimum expected standard for library branding on content web sites – support for institutional/customised logos, link text etc.
- Minimum expected mechanism for libraries to enable OpenURL linking – via online administration interface
- Documentation requirements

### 7.1.2. Quality Assurance Monitoring

For the code of practice to have teeth, independent and ongoing auditing will be required. The organisation responsible for developing the code should be responsible for initial audits with ongoing audits coming in a later phase of development (as has been the case with COUNTER).

A standard statistical measure for the accuracy of links from link resolver systems to publisher sites could be developed in a later phase. This could serve as a public record of link resolver effectiveness and help to raise overall performance/quality.

#### 7.1.3. Content Packages and Content Package Identifiers

The initiative should push for greater co-operation between publishers and link resolver suppliers over the issue of content packages, in particular the provision of data files to reflect bundled deals struck with national/regional consortia. This may, in addition, require pressure upon those negotiating these deals with publishers to demand collection data is supplied as part of the sales process.

The concept of common identifiers for content packages – between different link resolver solutions and across different library systems – should also be explored as part of the initiative.

#### 7.1.4. Journal Title Changes and Other Issues

Rules for a consistent approach to journal title changes should feed into the code of practice outlined above. In a similar vein, consistent approaches to combined issue and supplementary issue naming between content providers are badly needed to ensure accurate OpenURL generation and onward linking.

SIS does not believe it is the role of the organisation described here to own the development of these rule sets, operating as they do at a broader industry level than resolver knowledge bases alone. However, it is clear that such rules will impact on resolver knowledge bases and the code of practice described here. The organisation described could therefore participate along with others in the development of new, consistent approaches to these problems.

One possible route forwards might be an extension to the TRANSFER initiative to embrace these topics.

#### 7.1.5. A Code of Practice Cannot Guarantee Data Quality!

Of course, the code of practice outlined above cannot alone ensure that the data provided by content providers to link resolver suppliers is any more accurate. The ongoing role of educating publishers and others as to the need for accurate/current collection data and the benefits of this to all in the supply chain should not be under-estimated.

It is to be hoped however that the visibility achieved through the release of a code of practice will assist in this work. Nevertheless, a continuing emphasis on the need for poor quality data to be tackled at source (i.e. content providers) will remain necessary and will be a feature of the organisation's promotional activity.

## 7.2. ACCELERATING AND AUTOMATING KNOWLEDGE BASE DATA TRANSFER

In addition to the code of practice and other activities described above, which should seek to address the problems this study has identified in the short term, there are significant opportunities to explore the acceleration and automation of knowledge base data transfer between stakeholders, addressing data normalisation and timing issues further.

### 7.2.1. ONIX for Serials Online Holdings (ONIX SOH)

ONIX SOH is a member of the ONIX for Serials<sup>15</sup> family of XML formats for communicating information about serial products and subscription information, using the design principles and many of the elements defined in ONIX for Books<sup>16</sup>. Specifically, ONIX SOH is a format for communicating electronic serials holdings details from suppliers – such as hosting services, publication access management services, agents or publishers – to end customers in subscribing libraries. Other members of the ONIX for Serials family include Serials Products and Subscriptions (SPS) and Serials Release Notification (SRN).

A considerable number of interviewees, but especially the link resolver suppliers, see potential in ONIX SOH 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 localisation. With one common data format for all exchanges, the normalisation effort for the resolver suppliers would be reduced as far as 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 been slow, although some resolver suppliers report that they do receive ONIX SOH-formatted data in a limited number of cases. EDItEUR, the developer of the standard, certainly believes it to be a good format for describing the elements of a package of content, and perceives the main obstacle to successful rollout to be the general will to use ONIX SOH for this purpose.

Certainly, one challenge is the “don’t fix what isn’t broken” argument. A significant number of content providers, whose primary focus is on editorial and delivering content to the end user directly, believe they have addressed the needs of link resolvers with the data distribution arrangements they have already put in place. They may therefore have little incentive to move to a standard way of achieving the same outcome.

Furthermore, there are other concerns about the use of ONIX SOH in this way. There are some potential problems arising from the size and complexity of ONIX SOH messages (especially for the larger publishers with many titles in their collections). ONIX SOH adoption might also be a daunting task for many content providers. One resolver supplier argued that the potential for misunderstanding data in ONIX SOH format was considerable, because the data format is so complex.

EDItEUR states that it is aware of these concerns, but that a very rich message was given to it as the requirement. EDItEUR is also very open to

discussion about ONIX SOH and voiced a number of potential solutions that might ease its use: adoption of a more compact syntax; “flattening” of the code structure; implementing a profiled version of the standard expressly for knowledge base data transfers; implementing an update format (relaying only the *changes* to large and complex collection/holdings lists).

Conversely, some interviewees commented that – in one respect at least – ONIX SOH falls short of the immediate need. Whilst it provides a means to list web links for serial items (i.e. journal homepages) it does not offer a means to describe how to link to individual journal volumes, issues or articles within a given content package. Linking algorithm data is as critical to the effective operation of link resolvers as collection/holdings records. Presently, this crucial information would need to be separately transferred and there are no common approaches to either describing or relaying it.

#### 7.2.2. Web Services and ONIX SOH

In June 2006, EBSCO and OCLC reported the successful deployment of a web service to facilitate the automated delivery of electronic holdings data in ONIX SOH format to WorldCat<sup>17</sup>. Oliver Pesch, of EBSCO, stated that “Using a Web Service as the means to transfer holdings has great promise not only for updating WorldCat, but also for keeping the knowledge bases behind ERM systems and link resolvers up-to-date without the need for human intervention”.

The link resolver suppliers and some content providers recognise the potential of XML-based web services for automating (and therefore speeding significantly) data movement around the knowledge base supply chain. On the one hand, RSS could be utilised 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. On the other hand, and again drawing on a comparison to the usage statistics arena, a SUSHI<sup>18</sup> equivalent – a standard XML-based protocol for machine-to-machine harvesting of ONIX SOH-formatted data files – has potential for dramatically reducing the current drag in the timing of knowledge base updates. However, there is as yet no SUSHI equivalent, and whilst there has been talk of ONIX SOH-based web services, EDItEUR is not currently aware of any articulated demand to prioritise the further exploration of this.

#### 7.2.3. The need for experimentation

At the same time as addressing short term needs in the knowledge base arena via education and the code of practice outlined earlier, SIS recommends that stakeholders begin to explore the use of web services for knowledge base data transfers, most likely in conjunction with a stripped-down ONIX SOH file format.

Certainly, some early experiments should be attempted between content providers and link resolvers, although – as stated above – many publishers have put some kind of data distribution arrangement in place already and there may therefore be a more limited gain in the immediate term.

A more substantial gain is likely to result from experimentation between subscription agents and link resolver suppliers, as 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). There may be data protection and competition concerns from agents but SIS believes that one of the greatest opportunities in the existing supply chain is further co-operation between these two stakeholders.

Notes:

- ONIX SOH is, of course, intended to address the relay of holdings information for serials. As discussed earlier, the e-resource arena is fast moving to embrace other content types. Therefore, if the outcome of experiments is positive, there is likely to be demand for modifications to ONIX SOH (or extensions of it) to make the transfer of multi-genre collection and holdings data possible.
- A balance may need to be struck in addressing the data needs of link resolvers via ONIX SOH with those of other library systems (such as the library catalogue). From the content provider's perspective, there is considerable advantage in addressing any related supply chain issues at the same time and, if possible, with the same data output.

## 8. A CENTRALISED ALTERNATIVE

Currently, a number of link resolver suppliers independently collate data regarding online journal and book collections from information providers to create knowledge bases for their own products. We have referred to this proprietary approach as the “distributed” model.

As part of its research, SIS explored the concept of an alternative “centralised” approach, a model that would revolve around a single repository of content definitions and packages, underpinned by web services and automated publisher data processing using standard information formats. The metadata within this repository would be publicly-accessible to all that desired to use it (for example, the commercial link resolver suppliers, library home-grown solutions, and others).

This subject was deliberately discussed as a separate topic with stakeholders, to ensure that shorter-term recommendations for improving the existing supply chain would not be confused with an approach that, if it has merit, is clearly much more long-term.

### 8.1. REACTION TO THE CENTRALISED MODEL

Many publishers, other content hosts and librarians are attracted to a more centralised model of knowledge base creation and management.

For content providers, the benefit is clear: a single repository to send data to/communicate with, rather than the present model which requires arrangements with a number of different resolver organisations. There is an analogy with the bespoke bilateral reference linking agreements that preceded the establishment of the CrossRef linking framework in 2000.

When asked whether a centrally-managed knowledge base solution would be a welcome development, 62% of respondents to the online survey either agreed (34%) or strongly agreed (28%). Only 17% of respondents disagreed or strongly disagreed. Librarians see a number of benefits to a centralised solution. Chief amongst these is that the knowledge base information underpinning the resolver systems they acquire and operate would no longer be proprietary to their supplier. A number of librarians commented that they felt the situation they are presently in, reliant on a single supplier for a knowledge base that is tied to a linking service, was a risky, even unacceptable, one. In addition, a centralised knowledge base model would (they felt) move the competition from a focus on the knowledge base content itself to what different parties can actually do with the same underlying data, in terms of service and user interface etc. This shift in focus would be for the greater good of libraries and their users.

Furthermore, librarians feel that a public, central solution would have a greater visibility overall than the current model of numerous knowledge base suppliers, and that this ought to attract the attention of those content providers (especially the smaller players) who have yet to enter into the knowledge base supply chain.

However, many librarians also commented that a central solution would not or could not (due to scale) address a major area of work for them and (where relevant) their link resolver suppliers: the localisation of the details in knowledge bases to reflect individual, local conditions. That, in effect, a centralised solution could probably only realistically address the first piece of the puzzle: capturing the global collection details of all relevant content providers.

The level of interest expressed by the subscription agents interviewed in the idea of a centralised knowledge base system was varied.

For their part, link resolver suppliers stress the hurdles apparent in setting up a centralised solution. A prior attempt at creating such a model (the Jointly Administered Knowledge Environment, also known as JAKE) demonstrated, they argue, that centralisation may still not get the attention or involvement of the smaller players. In addition, if such a solution relies on voluntarily-supplied updates from content providers its knowledge base will slowly degrade in accuracy and value over time, as did indeed happen with JAKE. Furthermore, JAKE is a reminder that a central solution creates a single point of failure if it should ultimately be technically or economically unsuccessful.

The resolver suppliers also foresee issues with sourcing localisation data for a centralised solution (were this to be attempted). They cite the competitive issues in the market place that would limit the involvement of players that have library holdings data. Customised library collections are increasing, not lessening, and this problem may be more, rather than less, difficult to address with a monolithic, centralised solution than with the current model. A better focus, they believe, is on the refinement and automation of the existing, distributed framework.

Centralisation will also do little if anything, in of itself, to solve the quality assurance issues inherent in poor data from content providers. Any centralised solution would require a high level of data checking and maintenance from somebody (and who would do this?). Alternatively, it would demand that those harvesting data from the central pool understood its limitations and improved its quality, as happens already in the distributed model today.

A number of resolver suppliers were at pains to point out that the bulk of the competitive advantage in the present model lies chiefly in correcting/enhancing data from content providers, rather than (or in addition to) being the first to collect and load it, i.e. that a centralised model will not, in of itself, level the playing field amongst knowledge base suppliers or remove the proprietary nature of current solutions, as librarians might perhaps imagine.

## **8.2. WHO MIGHT LEAD AN INITIATIVE TO CREATE A CENTRALISED MODEL?**

The following organisations were all mentioned by interviewees as possible candidates for leading a centralising initiative:

- Bowker/Ulrichs
- CrossRef
- Google
- NISO
- OCLC
- Subscription agents

There was a strong feeling that the lead would need to be taken by an organisation that could demonstrate its independence but that also has strong links with all parties in the supply chain.

Of course, given that a number of the organisations listed above already own or operate proprietary link resolver solutions, there might be concerns from some quarters regarding their independence.

CrossRef already has a role as a metadata hub, and many we spoke to considered the knowledge base arena to be a logical extension of its current position in the market place. There might however be concerns from some that CrossRef is too publisher-centric, and a related issue could be that CrossRef itself might face a challenge in terms of introducing content providers other than primary publishers (i.e. other sources of the full text) into its model.

CrossRef does not see itself becoming a central knowledge base provider *per se*. Rather it is interested in becoming a channel for the distribution of its members' knowledge base-related data to the various link resolver suppliers. This certainly seems logical and, through automated, standards-based solutions, would provide tangible benefits to publishers and link resolver suppliers alike.

### 8.3. CONCLUSION

Many publishers, other content hosts and librarians are attracted to a more centralised model of knowledge base creation and management. They foresee communication and visibility benefits arising from a central system.

However, link resolver suppliers envisage difficulties in the operation of a completely centralised solution, and emphasise the significant task of data quality assurance (a task they currently undertake themselves). There is little enthusiasm amongst the link resolver suppliers for giving up this function, and certainly not whilst processes and standards to address data quality/consistency/comprehensiveness at source are still absent from the supply chain.

A centralised solution would require investment from some (as yet unknown) source or a charging model to operate successfully, and the quality assurance/management costs are likely to be substantial. Such a model would effectively be in direct competition with the existing link resolver systems and does not appear viable in the short-to-medium term. Without the quality assurance layer, the approach becomes more viable but the appeal for libraries is reduced.

The CrossRef organisation is interested in becoming a channel for the distribution of its members' knowledge base data to all of the existing link resolver systems. This is a potentially viable and valuable addition to the supply chain, worthy of exploration in the short-to-medium term.

## 9. AREAS FOR FURTHER RESEARCH

UKSG may wish to consider the following additional areas of potential research:

- Defining, in conjunction with experiments perhaps, a tailored version of the ONIX SOH data format for use in the knowledge base supply chain.
- Defining a SUSHI equivalent (XML harvesting protocol for collection/holdings files in ONIX SOH format), through consultation with parties in the knowledge base supply chain.
- Exploring standards-based mechanisms for describing and transporting inbound deep linking syntax details (not currently possible as part of ONIX SOH).
- Exploring the problems related to journal title changes and combined/supplementary issue naming. A code of practice would be a valuable outcome.

## 10. REFERENCES

*Web links listed below were accurate as of 25 January 2007.*

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](http://www.niso.org/standards/standard_detail.cfm?std_id=783)>
2. CrossRef. <<http://www.crossref.org/>>
3. The Digital Object Identifier (DOI) System. <<http://www.doi.org/>>
4. Caplan, P. and Flecker, D. "Choosing the Appropriate Copy – Report of a Discussion of Options for Selecting Among Multiple Copies of an Electronic Journal Article." Digital Library Federation Architecture Committee. September 1999. National Information Standards Organization (NISO). <<http://www.niso.org/news/reports/DLFarch.html>>
5. Van de Sompel, H. and Hochstenbach, P. "Reference Linking in a Hybrid Library Environment – Part 1: Frameworks for Linking." D-Lib Magazine 5.4 (1999). <[http://dx.doi.org/10.1045/april99-van\\_de\\_sompel-pt1](http://dx.doi.org/10.1045/april99-van_de_sompel-pt1)>
6. Van de Sompel, H. and Hochstenbach, P. "Reference Linking in a Hybrid Library Environment – Part 2: SFX, a Generic Linking Solution." D-Lib Magazine 5.4 (1999). <[http://dx.doi.org/10.1045/april99-van\\_de\\_sompel-pt2](http://dx.doi.org/10.1045/april99-van_de_sompel-pt2)>
7. Gold Rush – A Discovery and Management Tool for Electronic Resources. <<http://grweb.coalliance.org/>>
8. Chen, X. "Assessment of Full-Text Sources Used by Serials Management Systems, OpenURL Link Resolvers, and Imported E-Journal MARC Records." Online Information Review 28.6 (2004): 428-434. <<http://dx.doi.org/10.1108/14684520410570553>>
9. Wakimoto, J.C, Walker, D.S, and Dabbour, K.S. "The Myths and Realities of SFX in Academic Libraries". The Journal of Academic Librarianship 32.2 (2006): 127-136. <<http://dx.doi.org/10.1016/j.acalib.2005.12.008>>
10. United Kingdom Serials Group. <<http://www.uksg.org/>>
11. Koppel, T. "Solutions Forum: Visions of ERM." Managing Electronic Collections Workshop. National Information Standards Organization. September 2006. <<http://www.niso.org/presentations/MEC06-07-Koppel.pdf>>
12. UKSG: Transfer. <<http://www.uksg.org/transfer.asp>>
13. COUNTER – Counting Online Usage of NeTworked Electronic Resources. <<http://www.projectcounter.org/>>

14. "Z39.88-2004: The OpenURL Framework for Context-Sensitive Services, The Key/Encoded-Value (KEV) Format, Implementation Guidelines." National Information Standards Organization (NISO). December 2004.  
<[http://alcme.oclc.org/openurl/docs/implementation\\_guidelines/KEV\\_Guidelines-20041209.pdf](http://alcme.oclc.org/openurl/docs/implementation_guidelines/KEV_Guidelines-20041209.pdf)>
15. ONIX for Serials. EDItEUR/NISO.  
<<http://www.editeur.org/onixserials.html>>
16. ONIX for Books. EDItEUR. <<http://www.editeur.org/onix.html>>
17. "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](http://support.epnet.com/support_news/detail.php?id=250&t=h)>
18. NISO Standardized Usage Statistics Harvesting Initiative (SUSHI). National Information Standards Organization.  
<[http://www.niso.org/committees/SUSHI/SUSHI\\_comm.html](http://www.niso.org/committees/SUSHI/SUSHI_comm.html)>

## APPENDIX 1: INVITATION TO INTERVIEWEES

Dear <Name>

Scholarly Information Strategies (SIS – [www.scholinfo.com](http://www.scholinfo.com)) is conducting a study on behalf of the United Kingdom Serials Group (UKSG – [www.uksg.org](http://www.uksg.org)) into the new information supply chain that has developed to facilitate the creation of link resolver knowledge bases. This supply chain involves a number of parties: publishers and content hosts, subscription agents, libraries and providers of link resolver software tools. Given the significance of the knowledge base to the operation of OpenURL link resolvers, and by extension to the operation of libraries, the accuracy and timeliness of information being channelled through this supply chain is of critical importance.

Further background information regarding the study can be found here: [www.uksg.org/resolvers.asp](http://www.uksg.org/resolvers.asp).

I wondered if you (or a colleague) might be interested in sharing your views on this subject with us as part of this study. If this is of interest, we could look to set up a short phone discussion at a time convenient to you in October or November.

The results of this study will be presented at the annual UKSG conference in April 2007 in the UK, and will also be published in the UKSG's quarterly online journal *Serials*.

I hope you find this invitation to participate of interest and look forward to hearing from you.

<Sign Off>

## APPENDIX 2: INVITATION TO LIST SERVS

As readers of this list may know, Scholarly Information Strategies (SIS – [www.scholinfo.com](http://www.scholinfo.com)) is conducting a study on behalf of the United Kingdom Serials Group (UKSG – [www.uksg.org](http://www.uksg.org)) into the new supply chain that has developed to facilitate the creation of link resolver knowledge bases. This supply chain involves a number of stakeholders: publishers and content hosts, subscription agents, libraries and providers of link resolver software tools. Given the significance of the knowledge base to the operation of OpenURL link resolvers, and by extension to the operation of libraries, the accuracy and timeliness of information being channelled through this supply chain is obviously of critical importance.

Further background information regarding the study can be found here: [www.uksg.org/resolvers.asp](http://www.uksg.org/resolvers.asp).

SIS is approaching a cross-section of stakeholders to participate in the study through interview discussions. In addition to this methodology, we are interested in soliciting written comments from librarians subscribed to this list (to be sent to us off-list please). Specifically we are interested in any feedback on the following topics:

- What expectations do you have of the other parties in the supply chain (publishers/content hosts, subscription agents, link resolver suppliers)? What roles should they play, and what relationships are necessary to make the knowledge base supply chain operate effectively for you?
- What, in your opinion, is not working well in the present supply chain? What are the barriers to a smooth, timely, comprehensive and accurate knowledge base data flow that would ensure a good service to you and your users?
- How do you think the problems in the supply chain might be alleviated and who should take action to address them? The emphasis in this study is on simple, practical steps that parties in the chain can take in the short term to improve on the present environment.

We hope you find this invitation to participate of interest and we look forward to hearing from you off-list.

<Sign Off>

## APPENDIX 3: TEXT COPY OF ONLINE SURVEY

### Section 1: Introduction

This survey is being undertaken by [Scholarly Information Strategies](#) (SIS) on behalf of the [United Kingdom Serials Group](#) (UKSG).

We are investigating the issues and problems with the new supply chain that has developed to facilitate the creation of link resolver knowledge bases. Examples of link resolver knowledge base tools are Ex Libris *SFX*, EBSCO *LinkSource* or Serials Solutions *Article Linker*.

We have already conducted a number of telephone interviews with librarians and this online survey sets out to validate the findings from these discussions with a larger audience.

The survey will take approximately 10 minutes to complete.

We guarantee your anonymity in this survey. Neither SIS or UKSG will have any means of knowing which answers were received from which participants.

If you have any difficulties in completing this survey, please contact James Culling at SIS: [james@scholinfo.com](mailto:james@scholinfo.com).

Click on the **CONTINUE** button below to start.

**CONTINUE**

### Section 2: Screener Question

1. Does your organisation have a link resolver?

|     |                          |
|-----|--------------------------|
| Yes | <input type="checkbox"/> |
| No  | <input type="checkbox"/> |

*[If Yes, proceed to Section 3]*

*[If No, proceed to Section 7]*

### Section 3: About Your Link Resolver

2. Which link resolver system does your organisation use?

|   |                          |
|---|--------------------------|
| EBSCO <i>LinkSource</i>                 | <input type="checkbox"/> |
| Ex Libris <i>SFX</i>                    | <input type="checkbox"/> |
| Innovative <i>WebBridge</i>             | <input type="checkbox"/> |
| OCLC Openly Informatics <i>1cate</i>    | <input type="checkbox"/> |
| Ovid <i>LinkSolver</i>                  | <input type="checkbox"/> |
| Serials Solutions <i>Article Linker</i> | <input type="checkbox"/> |
| SIRSI <i>Resolver</i>                   | <input type="checkbox"/> |
| SwetsWise <i>Linker</i>                 | <input type="checkbox"/> |
| TDNet <i>TOUResolver</i>                | <input type="checkbox"/> |
| Other, please name:                     | <input type="checkbox"/> |

**Section 4: Problems with Link Resolver Knowledge Bases**

3. What percentage of publishers and content providers would you say are unaware of link resolvers and do not understand what they should do to work with these systems?

|               |                          |
|---------------|--------------------------|
| 1-10%         | <input type="checkbox"/> |
| 11-20%        | <input type="checkbox"/> |
| 21-30%        | <input type="checkbox"/> |
| 31-40%        | <input type="checkbox"/> |
| 41-50%        | <input type="checkbox"/> |
| More than 50% | <input type="checkbox"/> |

Comments: .....

4. What is your perception of the percentage of inaccurate publisher title/holdings data in your link resolver’s knowledge base?

|         |                          |
|---------|--------------------------|
| 1-5%    | <input type="checkbox"/> |
| 6-10%   | <input type="checkbox"/> |
| 11-25%  | <input type="checkbox"/> |
| 26-40%  | <input type="checkbox"/> |
| 41-60%  | <input type="checkbox"/> |
| 61-100% | <input type="checkbox"/> |

Comments: .....

5. Publishers are doing enough in managing journal title changes and journal transfers.

|                            |                          |
|----------------------------|--------------------------|
| Strongly agree             | <input type="checkbox"/> |
| Tend to agree              | <input type="checkbox"/> |
| Neither agree nor disagree | <input type="checkbox"/> |
| Tend to disagree           | <input type="checkbox"/> |
| Strongly disagree          | <input type="checkbox"/> |

Comments: .....

6. Some publishers and link resolver systems have difficulty in making title information for consortia deals available to libraries.

|                            |                          |
|----------------------------|--------------------------|
| Strongly agree             | <input type="checkbox"/> |
| Tend to agree              | <input type="checkbox"/> |
| Neither agree nor disagree | <input type="checkbox"/> |
| Tend to disagree           | <input type="checkbox"/> |
| Strongly disagree          | <input type="checkbox"/> |

Comments: .....

7. There is very little manual effort in configuring the link resolver's knowledge base to reflect local library holdings and conditions.

|                            |                          |
|----------------------------|--------------------------|
| Strongly agree             | <input type="checkbox"/> |
| Tend to agree              | <input type="checkbox"/> |
| Neither agree nor disagree | <input type="checkbox"/> |
| Tend to disagree           | <input type="checkbox"/> |
| Strongly disagree          | <input type="checkbox"/> |

*Comments:* .....

8. Subscription agents could assist libraries further by working more closely with link resolver suppliers to configure knowledge bases.

|                            |                          |
|----------------------------|--------------------------|
| Strongly agree             | <input type="checkbox"/> |
| Tend to agree              | <input type="checkbox"/> |
| Neither agree nor disagree | <input type="checkbox"/> |
| Tend to disagree           | <input type="checkbox"/> |
| Strongly disagree          | <input type="checkbox"/> |

*Comments:* .....

9. The effectiveness of link resolvers is hampered by the lack of a standard linking syntax for connecting to publisher web sites.

|                            |                          |
|----------------------------|--------------------------|
| Strongly agree             | <input type="checkbox"/> |
| Tend to agree              | <input type="checkbox"/> |
| Neither agree nor disagree | <input type="checkbox"/> |
| Tend to disagree           | <input type="checkbox"/> |
| Strongly disagree          | <input type="checkbox"/> |

*Comments:* .....

10. A significant problem for link resolvers is the generation of incomplete or inaccurate OpenURLs by databases (for example, Abstracting & Indexing products).

|                            |                          |
|----------------------------|--------------------------|
| Strongly agree             | <input type="checkbox"/> |
| Tend to agree              | <input type="checkbox"/> |
| Neither agree nor disagree | <input type="checkbox"/> |
| Tend to disagree           | <input type="checkbox"/> |
| Strongly disagree          | <input type="checkbox"/> |

Comments: .....

11. A centrally-managed knowledge base solution, with all content providers sending title/holdings data to one organisation only, would be a welcome development.

|                            |                          |
|----------------------------|--------------------------|
| Strongly agree             | <input type="checkbox"/> |
| Tend to agree              | <input type="checkbox"/> |
| Neither agree nor disagree | <input type="checkbox"/> |
| Tend to disagree           | <input type="checkbox"/> |
| Strongly disagree          | <input type="checkbox"/> |

Comments: .....

12. If you have any other comments on the problems and issues you experience with your link resolver's knowledge base, please enter them here:

13.If you have any other comments on how you think the problems in the knowledge base supply chain can be solved, please enter them here:

### Section 5: Demographic Questions

14.What type of library do you work in?

|  |                          |
|--|--------------------------|
| Academic institution with research programme | <input type="checkbox"/> |
| Other academic                               | <input type="checkbox"/> |
| Corporate                                    | <input type="checkbox"/> |
| Hospital                                     | <input type="checkbox"/> |
| Government                                   | <input type="checkbox"/> |
| Medical library                              | <input type="checkbox"/> |
| Medical library associated to a university   | <input type="checkbox"/> |
| Public                                       | <input type="checkbox"/> |
| Some other type of library                   | <input type="checkbox"/> |

15.Where do you work?

|                           |                          |
|---------------------------|--------------------------|
| Africa                    | <input type="checkbox"/> |
| Asia                      | <input type="checkbox"/> |
| Australasia               | <input type="checkbox"/> |
| Central and South America | <input type="checkbox"/> |
| Eastern Europe            | <input type="checkbox"/> |
| Middle East               | <input type="checkbox"/> |

|                |                          |
|----------------|--------------------------|
| North America  | <input type="checkbox"/> |
| UK             | <input type="checkbox"/> |
| Western Europe | <input type="checkbox"/> |

### Section 6: Thank You and Follow Up

Thank you for taking the time to complete this survey. Your help is very much appreciated.

The results of this study will be presented at the UKSG annual conference in April 2007 and a paper will also appear in the UKSG's quarterly online journal *Serials*.

If you would be interested in discussing this topic further with us via a telephone interview, please provide your email address and phone number below so that we may contact you:

|                |  |
|----------------|--|
| Email address: |  |
| Phone number:  |  |

### Section 7: Thank You and Exit

We are looking for feedback from organisations that are currently using a link resolver. Your answer indicates that you do not have a link resolver. Thank you for your interest in this survey.

## **APPENDIX 4: ABOUT SCHOLARLY INFORMATION STRATEGIES**

### **CHRIS BECKETT**

Chris Beckett has over 25 years experience as a manager and provider of information products and services. He founded his first consultancy in 2002 to help publishers, intermediaries and libraries develop and deliver electronic products and services. He has considerable senior executive experience in product development, sales and marketing for industry intermediaries including Vice-President Library Services for Ingenta, International Publisher Sales Director for Ingenta, and Sales and Marketing Director of CatchWord. Prior to CatchWord, Chris held a number of senior management positions at Blackwell's Information Services including General Manager of Blackwell's Online Bookshop, which he founded, and was a member of Blackwell Information Services Executive Board. At Blackwell's he was responsible for initiating a number of strategic business developments including, Blackwell's first online service (Connect), their consolidation service, and the Uncover Document Delivery service (jointly with the UnCover Company).

### **SIMON INGER**

Simon Inger has been working in the journals industry for over 18 years. In this time he has worked for B.H.Blackwell, CatchWord, Ingenta and, since March 2002, as an independent consultant. Simon was co-founder and Managing Director of CatchWord Ltd, the world's largest e-journal hosting organization, from its inception in 1995 to its sale to Ingenta Plc in February 2001. Simon has always seen himself as the interface between the technical and non-technical minds within the industry but his interests and expertise span broad topics such as reference-linking, archiving, access control, information standards, purchasing models, internet infrastructure and rights management. Simon is often invited to speak at library and publishing conferences on subjects from global e-journal delivery to library portals, OpenURL and linking and their impacts on purchasing and consumption.

### **TRACY GARDNER**

Tracy Gardner has more than 7 years experience in marketing communications. Most recently, she served as Marketing Manager for CABI, one of the leading publishers of bibliographic databases, books, CD-ROMs and Internet resources in the applied life sciences. Previously, at CatchWord, Tracy was in charge of marketing communications and market research for the, then largest, electronic journal hosting company, and gained considerable experience developing marketing services for electronic journal content aggregators.

### **GEOFFREY BILDER**

Geoffrey Bilder has over 15 years experience as a technical leader in scholarly technology. He co-founded Brown University's Scholarly Technology Group in 1993 to provide advanced technology consulting on issues related to academic research, teaching and scholarly communication. He later served as head of R&D in the IT department of Monitor Group, a management consulting firm based in Cambridge MA. Most recently, Geoffrey was Chief Technology Officer at Ingenta, and he has spoken extensively to publishers and librarians on how emerging social software technologies are likely to affect scholarly and professional researchers.

### **JAMES CULLING**

James Culling has considerable experience and expertise in the areas of web publishing and library systems, having worked with scholarly publishers and academic libraries for 10 years. His experience is varied, ranging from overseeing the delivery of customised websites and content hosting services for a range of international publishers, to managing the installation and support of link resolver and portal products for libraries. He has a number of years experience as a software and technology trainer and has also been involved in e-journal pricing and licensing negotiations between publishers and library consortia. He most recently worked for library systems vendor Ex Libris, and has also served at online publishing service providers Ingenta and CatchWord, and the subscription agent Swets Blackwell.