



| Project Information | | | |
|--|---|-----------------|----------------|
| Project Acronym | - | | |
| Project Title | A low-cost OpenURL resolver for UK HE/FE: a scoping study | | |
| Start Date | February 2008 | End Date | September 2008 |
| Lead Institution | EDINA, University of Edinburgh | | |
| Project Director | Christine Rees | | |
| Project Manager & contact details | Leah Halliday Leah.halliday@ed.ac.uk | | |
| Partner Institutions | - | | |
| Project Web URL | http://edina.ac.uk/projects/LowCostOpenURL_summary.html | | |
| Programme Name (& number) | | | |
| Programme Manager | Tom Franklin/ James Farnhill | | |

| Document Name | | | |
|-------------------------------------|---|---|--|
| Document Title | Maintaining an OpenURL resolver knowledge base through record sharing (Addendum to Final Scoping Report) | | |
| Author(s) & project role | Leah Halliday, Tim Stickland and Christine Rees | | |
| Date | | Filename | |
| URL | http://www.edina.ac.uk/projects/openurladdendum.pdf | | |
| Access | <input type="checkbox"/> Project and JISC internal | <input checked="" type="checkbox"/> General dissemination | |

| Document History | | |
|------------------|--------------|--------------------------|
| Version | Date | Comments |
| V0a | 2 March 2009 | Draft submitted to JISC. |
| V1 | March 2009 | Final version |
| | | |

Maintaining an OpenURL resolver knowledge base through record sharing

On submission of the project report to scope a low-cost OpenURL resolver for UK HE/FE¹, JISC indicated that as demand for a resolver based on a centralised Knowledge Base (KB) is likely to be low and commercial offerings are readily available, the cost of such a development is not justified. It suggested that EDINA explore the possible benefits of record sharing between libraries that have already purchased an OpenURL resolver² as a partial solution to the problems outlined in this report. This short report reflects this request.

Background

As indicated in the scoping report, although publishers and aggregators (content providers) sell bundles of journals to libraries, the contents of these bundles are not standard, i.e. they vary depending on the negotiated deal. Thus, a bundle negotiated for UK HE by NESLi2³ contains a sub-set of the total package available from that provider (across the globe). In fact, a large bundle may contain several subject-based sub-bundles each of which, within the UK, may be a slight variation on the standard 'global' subject bundle. The content of a bundle may also change from time to time due to changes in ownership. Titles are sold by one publisher to another and thus disappear from one bundle and re-appear in another. Until the resolver is configured to reflect these changes, these titles are rendered invisible to users. (While core, popular titles are usually stable, the collection of smaller, specialised titles varies across time).

The content of a bundle is not the only variant. Access is often negotiated from a specific date, e.g. from 1997 onwards so this may be the 'base' date for a NESLi2 deal. Some libraries, that began their subscriptions much earlier, however, may be entitled to access earlier volumes. Furthermore, the backfile may be available as a separate subscription. While some of this information is library-specific, much of it is UK-specific. Every variant to the bundle as it appears in the KB must be maintained by the library as a local exception. Understandably, libraries wish to minimise the volume of exceptions. Where an exception applies to the whole UK market, libraries would like the KB vendor to configure it rather than having to configure it on a library-by-library basis.

One of the biggest problems for libraries using a KB-based Open-URL resolver is the lack of *accurate* and *current* information about 'big deals' such as the NESLI2 licences (which are UK specific).

Configuration and maintenance of KBs to reflect the 'local' profile of 'big deals' is resource intensive and duplicates effort. This is one of the main complaints of libraries about OpenURL resolvers. Libraries want to know the current contents of the bundles to which they are subscribed but, at present, this information is not

¹ <http://www.edina.ac.uk/projects/openurlreportfinal.pdf>

² 92 UK academic institutions are registered with the OpenURL router <http://openurl.ac.uk/doc/>.

³ <http://www.nesli2.ac.uk/> - the UK's national initiative for the licensing of electronic journals on behalf of the higher and further education and research communities

provided to them. Content Complete, the negotiating agent for NESLI2 considers it the responsibility of the content provider and the KB vendor to ensure that this information is accurately represented in the KB. Clearly, the authoritative source of information about the content of a bundle is the content provider and although some publishers currently provide this information to KB vendors, many do not. Libraries can obtain information about the content to which they are entitled access directly from the publisher but this information then has to be imported manually into the KB and maintained every time the KB is updated. This is resource intensive. Libraries want to minimise the requirement for local amendment; they want the information about bundles to be configured in the KB. This requires content providers to deliver the information in a timely manner and in a useable format to KB vendors. As only a few currently do this, libraries have a problem and some work together to address it. Often, libraries rely on one another to identify changes. A library discovering that a title has been removed from a bundle or has moved from one bundle to another will alert other libraries through a relevant mailing list (such as the KB vendor user group or the JISC Mail list *lis-e-resources*). Libraries work together on the mailing lists to try to arrive at consensus on what is and is not included in a bundle.

In our scoping report, we suggested that although it may not be possible to develop a sufficiently functional OpenURL resolver at 'low cost' it may be possible to reduce the costs to individual libraries within the community by sharing the cost of configuring the KB to reflect UK-based licence deals such as those negotiated by NESLI2.

Sharing Records

Sharing of bibliographic records between libraries is beneficial. Records for electronic serials are expensive to create or buy and sharing of records spreads that cost. Records sharing is also an established practice. Many libraries share records through OCLC's WorldCat. Within the UK, sharing of records and of updates is facilitated by SUNCAT, the Serials Union Catalogue.

While records sharing is a useful and established activity, unfortunately, it does not solve the problem highlighted in this report, namely that:

- a KB must be configured to reflect the collection of the subscribing library;
- this is a resource-intensive activity;
- for large bundles of journals this must be done manually because neither the content provider nor the negotiating agent can provide details of what the bundle contains in a form that the library can use to configure the database;
- the content of bundles can change from month to month and thus, the configuration of the KB must be maintained.

In brief, the problem lies not with record creation but with maintaining a database that accurately reflects the current content of e-serials bundles for the UK and whose contents change. This problem relates to data about 'holdings' rather than to bibliographic data.

It may be useful for libraries using the same system, e.g. all those using SFX, to collaborate to configure the KB for that system so that it reflects UK-based licences. Our interviewees told us that librarians using the same system are already in contact with one another through the system user group and assist one another with set-up and

implementation. Collaboration to assist one another with configuration of the KB would be greatly facilitated by a central initiative.

Since collaboration at this level involves sharing data between KB products, two approaches are possible. Firstly, a framework could be established for standardizing the representation of journals, bundles and subscriptions from different KB products (possibly based around ONIX for Serials), and vendors could be encouraged to support this framework by enabling export and import of standardized data from their KB. Secondly, librarians could be encouraged to share data using existing KB features, accepting that sharing between different products would be limited.

The first option has the potential to be a project of significant scale and complex scope, and would require considerable "buy in" from vendors. We believe the second option is far more realistic, since it would be lightweight and its success is primarily dependent on encouraging activities that librarians already seem to be keen to engage in.

A central initiative enabling sharing KB data could be based around an open source repository product (ePrints, DSpace and Fedora are likely candidates), which should not require extensive development work. The content would be user contributed data, deposited for the use of other librarians at their own discretion, so particular representations of (for example) journal bundles would not be presented as having any central authority. This approach is lightweight: by placing the focus on collaboration and sharing amongst the community however, procedures for validation and verification would also be the responsibility of the community⁴.

Certain aspects of the scope of the work to support a central initiative and dependencies, e.g. constraints arising from proprietary rights of the software vendor, were unknown to us. As an initial step, in order to gain some understanding of the current state of play and the feasibility of the proposed initiative, we consulted the Serials librarian from a large University using the market-leading resolver and a representative of the vendor of that system. Clearly, this consultation exercise was not representative but it was sufficient to extend our understanding of the field.

Both of the parties consulted are familiar with and are currently wrestling with the problem. Both would welcome and would participate in a central initiative to facilitate sharing of information about the current content of 'big deals' such as the NESLI2 deals. The vendor would welcome a central initiative benefiting the whole community or an initiative that would address the problem for her users, e.g. whereby one user provides a current configuration and the vendor facilitates sharing of that configuration across the user group. The librarian considers it essential that a central initiative should benefit the whole community.

It seems that if a central initiative were to facilitate creation of a file in a standard format containing the journal title, identifier (e.g. the ISSN), start year, volume, issue, and, if applicable, end year, volume and issue, the vendor would create an autoloader

⁴ This contrasts with services such as SUNCAT, which act as authoritative sources of journal information, and rightly require "heavyweight" procedures

that downloads the data, inserts them into the KB and creates reports for inspection and rectification of any errors if necessary.

One issue that would not be easily addressed is ‘validation’ of the data. The vendor told us that it is impossible to validate the data without the co-operation of the libraries but the librarian explained that libraries simply do not know whether a list of the titles that are allegedly contained in a bundle is valid or not. Librarians consider the content providers (publishers or aggregators) to be the authoritative source of that information and as the content providers will not provide the required data, establishing the authority of files in a Central database would be a challenge. She explained that the user group of which she is a member does use the mailing list to arrive at consensus regarding the content of specific bundles but this is a pragmatic solution and she would have concerns about extending it to the whole of the UK sector.

The solution being proposed here is based on ‘unvalidated’ data. We asked the librarian whether she would check data downloaded from a central file before using it. She would do so in the first instance. She said that, with time, she would probably come to trust the source but in the first instance she would wish to check the data in any downloaded file before using it to configure the information provided to her users. This suggests that a central facility offering ‘unvalidated data’ may be useful. It may be necessary for librarians to invest time, in the first instance, to reassure themselves regarding the validity of data but as a means of sharing what they know, the facility may be valuable to them. The librarian interviewed explained that it would be important to include a facility for commenting on each revision so that participants could see, at a glance, whether the ‘alerted’ change was known or new to them.

When asked if she would impose any conditions on participation, the Librarian explained that while her library would impose no conditions, she would wish publishers to have access and the facility to amend the central database, and would anticipate that publishers would require some level of confidentiality with regard to their own data, i.e. whilst making the data available to libraries and KB vendors, they would not necessarily wish to allow other publishers to access them.

Recommendation

Our consultations with a librarian and vendor suggest that it would be worthwhile scoping the development of a central initiative to facilitate sharing the effort of configuring KBs, within UK research libraries, to reflect the UK-specific contents of ‘big deals’.