

Discovery and access: publisher–library collaboration on standards

Library users want to discover and get access to relevant publications easily. In the UK, the Joint Information Systems Committee (JISC) is developing an Information Environment that will enable more seamless discovery and access for the academic community, based on standards and protocols that allow online services to work together and become more interoperable. The PALS Metadata and Interoperability programmes were funded by the JISC to encourage publishers and libraries to work together using these standards and develop joint solutions for improving interoperability throughout the information chain. The second PALS programme ended in 2006 and demonstrated how the use of standards can improve discovery for users, improve electronic resource management for libraries, and improve the visibility of electronic publications. This article summarizes what the PALS 2 programme and its projects achieved.



CHRISTINE BALDWIN

Consultant

Information Design & Management

Background

Library users want to discover and get access to relevant publications easily. Enabling this is a challenge considering the wide range of online services available, each with its own interface, features and search facilities. In the UK, the Joint Information Systems Committee (JISC)¹ is developing an Information Environment (IE)² that will enable more seamless discovery and access for the academic community. The IE is based on using standards, protocols and techniques that allow online services to ‘work together’ in a secure way. The many online services and publications remain separate, but information about them (metadata) is exchanged across systems.

This ‘interoperability’ of metadata across systems is a key principle of the IE. The problem is how to achieve it in a practical way across different publisher and library systems. When the JISC started planning the IE, it brought the problem to PALS (Publisher and Library/Learning Solutions). PALS was a forum that brought UK publishers (ALPSP and the Publishers Association) and the academic community (JISC) together to develop joint solutions to electronic publishing issues. PALS agreed that interoperability was an important issue, and solutions could lead to benefits on both sides – more seamless discovery

and access for users, increased visibility of electronic publications and, in principle, greater use.

In 2002, PALS held a conference to introduce publishers to the IE, and a ‘5 step guide’³ was published. Some of the guidelines are as applicable today as they were in 2002, for example: expose metadata about your publications for distributed searching and harvesting; share news and alerts using RSS; implement OpenURLs, and use persistent URIs. PALS also set up the PALS Metadata and Interoperability working group⁴ to plan some practical work. The group reasoned that a good starting point was to get publishers and libraries working together using the standards that would enable interoperability. As a result, the JISC established the PALS Metadata and Interoperability programmes.

The PALS programmes

The PALS Metadata and Interoperability programmes are JISC-funded projects where publishers work with the academic community to use interoperability standards, tackle implementation issues, and explore the potential benefits. The first programme (PALS 1)⁵ was funded in 2003 and comprised

six short projects. They focused on IE standards such as OAI-PMH, RSS and DOIs and resulted in a range of reports, case-studies and demonstrators.

The second programme (PALS 2)⁶ ended in 2006, and the eight projects are discussed in this article. Like the first programme, each project was short and involved collaboration between publishers and the academic community on metadata and interoperability standards. However, PALS 2 went beyond simply ‘exploring the issues’. The focus was on using the standards in innovative ways, and creating practical tools and guidelines for the community that would make adopting the standards easier. The range of standards used was also broader, including OAI-PMH, RSS, SRU, OpenURL, publisher initiated standards like ONIX for Serials and ONIX for Licensing Terms, and jointly developed standards like COUNTER.

This article focuses on the interoperability ‘problem’ each addressed, the solution they devised, and the significance of the work. Each project’s deliverables and final report (with full technical details of the work) can be found on its project website.

Electronic expression of licensing terms

Libraries sign licence agreements for a wide range of digital resources, and the terms contained in these agreements can vary widely. In order for libraries to comply with the licences, they need to be able to communicate the licence terms to their users, so users know what they can do when they use the resources. This is difficult if paper licence agreements are stored in filing cabinets. What is needed is the ability to express licence terms electronically so that the terms are actionable.

EDItEUR has developed the ONIX standards, a family of XML formats for communicating rich metadata about books, serials, and other published media, using common data elements. EDItEUR is now developing ONIX for Licensing Terms (OLT)⁷, XML formats for expressing and communicating licence terms, building on the work of the Digital Library Federation’s Electronic Resource Management Initiative (ERMI)⁸ and the joint EDItEUR/NISO work on ONIX for Serials⁹.

It is envisaged that OLT will work in the following way. A given licence agreed by a publisher and library is expressed in ONIX Publications Licence format (ONIX-PL), and the licence terms are

defined in a data dictionary. The licence is sent to the library’s Electronic Resource Management (ERM) system as an XML message. The ERM system looks after user authentication and links the actionable licence terms to the relevant resources. When users access the resources, they are then informed about the usage terms.

The PALS 2 projects on OLT focused on how to implement OLT in a practical way. The first project – XML Expression of a Publisher/Library Licence¹⁰ – involved mapping the Wiley InterScience Enhanced Access Licence for Academic Customers into ONIX-PL format. The project involved close collaboration between BIC, Wiley and Cranfield University to reach agreement on which licence terms should be actionable and their precise semantic meaning. It resulted in a detailed specification for the ONIX-PL format, the first release of the OLT Dictionary, and an ONIX-PL expression of the Wiley licence. More terms will be added to the OLT Dictionary as further licences are mapped and new OLT formats are developed.

For OLT to be used on a wide scale, tools will be needed to facilitate mapping licences into ONIX-PL. BIC, ALPSP and Loughborough University collaborated on the second PALS 2 project – Specifying Publisher Tools and Library Benefits¹¹ – to gather requirements for the tools and explore the potential benefits with publishers and libraries. They found that libraries were as interested in having tools as publishers. Having licences in electronic format will not only improve compliance, but could help libraries to compare licence terms and negotiate licences. The project resulted in a specification for the drafting tools that will allow both publishers and libraries to create ONIX-PL licences using templates.

EDItEUR is continuing the work of the PALS 2 projects in follow-up extensions. The JISC, Publishers Licensing Society (PLS), and EDItEUR are funding development of the drafting tools. The JISC is funding development of the first template for the tools, the JISC model licences. EDItEUR is also working with vendors of ERM systems to facilitate the implementation of OLT in libraries.

Improving usage statistics

COUNTER¹² is an international initiative to improve the reliability and comparability of usage statistics for online publications. COUNTER’s

Codes of Practice set standards for the recording, reporting, and delivery of usage statistics from vendors to libraries. COUNTER is always seeking to improve the quality of the statistics generated according to its Codes, and this was the starting point for its PALS 2 project, COUNTER Filter¹³. An independent study by Davis and Price¹⁴ suggested that the design of a vendor's electronic interface could have a measurable effect on the usage statistics generated. For example, if a vendor requires users to view an HTML version before viewing the PDF, this could inflate the number of full-text downloads. COUNTER felt that it would be useful to investigate this potential problem and propose solutions.

During the project, COUNTER worked with publishers to develop and test two data filters. The 'unwanted HTML' filter was not viable, but the 'unique article' filter was. It compensates for the inflation of usage statistics by providing a new metric – the number of successful unique article requests in a session. COUNTER will recommend that vendors use the unique article filter and that the new metric is included in the next Code of Practice. The project also conducted a survey of current vendor practice on implementing unique article identifiers. As there was great variation in practice, they have recommended some best practice guidelines to include in the next Code of Practice.

Social bookmarking

Social bookmarking services like del.icio.us¹⁵ allow anyone to create a personal collection of links, organize that collection using 'tags', and share their tags with other users. Nature Publishing Group (NPG) became interested in the potential of social bookmarking to improve discovery and developed an experimental service for academic users called Connotea¹⁶. As NPG actively encourages its authors to deposit their articles in institutional repositories, it wanted to explore how repositories could be integrated with social bookmarking services. Integrating these two environments has the potential to improve discovery and provide a more seamless experience for users.

NPG's PALS 2 project, Dictate (A Distributed Content Tagging Tool for EPrints)¹⁷, focused on this integration. They developed an open source Tagging Tool¹⁸ that enables repositories running

EPrints software¹⁹ to integrate with del.icio.us, Connotea, or any social bookmarking service running the open source Connotea code. Once the administrator of an EPrints repository has installed the Tagging Tool, users can take advantage of social bookmarking features without leaving the repository environment. They can see what tags other users have assigned to the article they are viewing in the repository, create a bookmark for the article in a social bookmarking service, and follow links to articles with the same tags as suggested by the service.

The NPG team feel that integrating institutional repositories with social bookmarking services will have real benefits for the academic community. Users will have a more seamless discovery environment. Shared tagging should enable new forms of navigation and discovery within repositories, facilitate linking to the external literature, and increase the visibility of institutional repositories.

Disclosing journals using OAI

OAI-PMH²⁰ is a protocol used widely in the academic community for disclosing metadata about publications. It specifies how to make metadata available for harvesting in a standard way, so that service providers can aggregate metadata from diverse sources for cross-searching. A project in the first PALS programme²¹ showed how publishers can use OAI-PMH to disclose their journals. However, setting up an OAI-compliant repository does require some technical expertise, and this presents a barrier to the small publishers that would benefit most from the disclosure.

The Centre for Digital Library Research (CDLR) at the University of Strathclyde set out to develop a low-tech solution to OAI-based disclosure for small publishers. Their PALS 2 project, STARGATE (Static Repository Gateway and Toolkit)²², was based on the 'static repositories' model²³ for using OAI-PMH. Instead of building an OAI-compliant repository, a data provider builds a static repository (SR), effectively an XML file of the relevant metadata on an accessible server. A separate static repository gateway handles the technical aspects of making the metadata available for harvesting, i.e. the complexity is shifted away from the publisher. During the project, CDLR worked with four library and information science journals, set

up static repositories for each one, and used an SR gateway to enable disclosure via services like TechXtra, OAIster, and METALIS.

STARGATE demonstrated that static repositories are easy to create, the technology works, and the metadata disclosed via the SR gateway is interoperable. The project created a range of tools and guidelines that will allow even the smallest publisher to use the SR model to disclose their journals to the academic community. However, for the approach to be used on a wide scale, a permanent SR gateway for publishers will be needed. The JISC has granted CDLR a short project extension to document how to set up and run such a gateway for publishers.

Disclosing journals in library catalogues

Though much of an academic library's budget is spent on acquiring serials, most Online Public Access Catalogues (OPACs) contain only records for the journal titles, not the articles they contain. Emerald Group Publishing reasoned that if it were possible to add data about journal articles to the library OPAC, this could improve their visibility and discoverability, and deliver a more integrated OPAC experience to library users.

Emerald collaborated with Talis and the University of Derby on a PALS 2 project called TOCRoSS (Table of Contents by Really Simple Syndication)²⁴ to see if RSS could be used to place journal table of contents (TOC) data into a library OPAC without human intervention. RSS is a standard for transmitting news feeds, and many publishers use it for TOC alerting services aimed at end-users. Using RSS to transmit feeds to a library OPAC was therefore an innovative use of the standard.

They extended the RSS 2.0²⁵ specification to carry metadata about journal articles in ONIX for Serials SRN (Serials Release Notification)²⁶ format. Talis then developed the software for handling the TOCRoSS feeds and generating MARC records for each article. During the project, TOC data for 160 Emerald journals (3,000 articles) was imported into the OPAC at the University of Derby. Users were able to search using keywords, retrieve journal article records, and view the full text, and overall feedback was positive about including article records in the OPAC.

Emerald has developed a 'Publisher Starter Kit'²⁷ with associated software to assist other content providers implement TOCRoSS feeds. TOCRoSS was conceived as a standards-based technology for enabling Web 2.0 applications, and successfully demonstrated this by importing TOC data into a library OPAC. There is potential to use it for other applications, for example an alerting service from a single service provider using metadata from many publishers.

Updating journal holdings in library catalogues

SUNCAT²⁸ is the national union catalogue of serials held by UK research libraries and is based at the EDINA data centre at the University of Edinburgh. Keeping SUNCAT up to date with electronic serials data for over 60 libraries is a challenge. Licensing agreements are complex, the serials covered by them may change, and data is received from many sources in a range of formats. EDINA reasoned it would be more efficient if publishers, aggregators, and libraries all transmitted holdings data in a standard format.

ONIX for Serials SOH (Serials Online Holdings)²⁹, one of the ONIX for Serials family of standards, is a format designed for just that purpose. EDINA and Serials Solutions collaborated on project AIMSS (Automating Ingest of Metadata on Serials Subscriptions)³⁰ to use SOH to update the holdings data in SUNCAT. Serials Solutions is a publication access management service (PAMS) and has a knowledge base of holdings data for its library customers, many of which are SUNCAT libraries. During the project, Serials Solutions created SOH messages with holdings data for two customers and transmitted it to EDINA where it was parsed, extracted, and converted to MARC21 format, mapping SOH to the MARC 856 tag. Full details are in their final report and a recent article in *Serials*³¹.

AIMSS demonstrated that SOH can be used to update serials holdings in library catalogues, and that it is relatively easy to develop a capability to create and process the messages. As the standard is fit for purpose, they argue that more take-up is desirable. In principle there are many scenarios for using SOH involving PAMS, aggregators,

publishers and libraries. Transmission from PAMS to library (or union catalogue) would seem to have advantages, as a PAMS has holdings data for many libraries.

Web services for resource discovery

Web services have the potential to improve resource discovery, enabling separate digital libraries, portals, and virtual learning environments (VLEs) to be more open and interoperable. The National e-Science Centre (NeSC) was developing training and digital library environments for two large EU projects on grid computing³² and wanted to use the service-oriented approach for resource discovery. They needed to develop ‘machine services’ that would work behind the scenes to aggregate and repurpose metadata from many sources and present them to users in environments that met their needs.

NeSC collaborated with Edinburgh University Library on a PALS 2 project called metadata+ (Machine Services for Metadata Discovery and Aggregation)³³ to develop a test bed of machine services for metadata discovery and aggregation. The core of the test bed was a Fedora repository containing 15,000 metadata records with machine services for searching (SRU) and linking (OpenURL). A demonstrator on the project’s website shows how the SRU protocol can be used for discovering and aggregating metadata (local and remote) and repurposing it for different contexts. For example, it can be mapped to MODS for use in a digital library, or to IEEE LOM for use in a VLE. It also demonstrates a novel use of the SRU protocol to map the metadata to containers for export, e.g. to dynamically create e-learning content packages (SCORM), digital library metadata collections (MODS collection), and news feeds (RSS).

The project demonstrates how publisher metadata can be aggregated from multiple sources and then presented to users in different contexts. Several digital library initiatives are using the test bed infrastructure in innovative ways, for example using the SRU RSS service for podcasting and alerting for new content, and wiki approaches for content management that allow users to enrich metadata with reviews or recommendations.

Outcomes

The PALS Metadata and Interoperability programmes have demonstrated that publishers and libraries can work together and develop joint solutions for improving interoperability throughout the information chain. The PALS 2 programme has resulted in practical outputs such as tools and guidelines for the community to use, and practical experience using the key standards and protocols effectively.

The PALS 2 work also illustrates that standards are a means to an end. Many standards exist, and both publishers and libraries need to make choices about the standards they will use. Each project identified an area where using standards was likely to result in real benefits. Together they demonstrate that interoperability can:

- improve disclosure and discovery
- provide a more integrated user environment and seamless user experience
- improve electronic resource management
- increase the visibility of electronic publications
- contribute to an evidence base for decision making.

Understanding the benefits will be important in stimulating the wider take-up of standards. In December 2006, the JISC held a seminar, *Discovery and Access: Standards and the Information Chain*³⁴, to consult with the community about the future. The seminar reviewed interoperability standards from publisher, library, and other stakeholder perspectives, identified gaps and issues that need to be addressed, and highlighted areas for future work. Presentations given are posted on the seminar website, along with a summary of the seminar and discussion. The areas for future work may lead to a PALS 3 programme.

For those who want to know more about the PALS Metadata and Interoperability programmes, there is a synthesis³⁵ of the work to date on the JISC website. This gives more detail on what the programmes have achieved and what the individual projects have done. It provides links to the many useful outputs created, plus links to useful resources on interoperability standards generally.

References

1. Joint Information Systems Committee (JISC):
<http://www.jisc.ac.uk/>
2. JISC Information Environment:
http://www.jisc.ac.uk/whatwedo/themes/information_environment.aspx
3. Powell, A, 5 step guide to becoming a content provider in the JISC Information Environment, *Ariadne*, 2002, 33.
<http://www.ariadne.ac.uk/issue33/info-environment/>
4. PALS Metadata and Interoperability working group:
http://www.jisc.ac.uk/aboutus/committees/working_groups/pals_interop_metadata_group.aspx
5. JISC PALS 1 programme:
http://www.jisc.ac.uk/whatwedo/programmes/programme_pals.aspx
6. JISC PALS 2 programme:
http://www.jisc.ac.uk/whatwedo/programmes/programme_pals2.aspx
7. ONIX for Licensing Terms:
http://www.editeur.org/onix_licensing.html
8. Digital Library Federation’s Electronic Resource Management Initiative (ERMI):
<http://www.diglib.org/standards/dlf-erm02.htm>
9. ONIX for Serials:
<http://www.editeur.org/onixserials.html>
10. Electronic Expression of Licensing Terms: XML Expression of a Publisher/Library Licence:
http://www.jisc.ac.uk/whatwedo/programmes/programme_pals2/project_onixlt1.aspx
11. Electronic Expression of Licensing Terms: Specifying Publisher Tools and Library Benefits:
http://www.jisc.ac.uk/whatwedo/programmes/programme_pals2/project_onixlt2.aspx
12. COUNTER:
<http://www.projectcounter.org/>
13. COUNTER Filter (Improving the comparability of usage statistics and the implementation of unique article identifiers):
http://www.jisc.ac.uk/whatwedo/programmes/programme_pals2/project_counter.aspx
14. Davis, P M and Price, J S, eJournal interface can influence usage statistics: implications for libraries, publishers, and Project COUNTER, *JASIST*, 2006, 57(9).
<http://people.cornell.edu/pages/pmd8/interface.doc> (Final manuscript)
15. del.icio.us:
<http://del.icio.us/>
16. Connotea:
<http://www.connotea.org/>
17. Dictate (A Distributed Content Tagging Tool for EPrints):
http://www.jisc.ac.uk/whatwedo/programmes/programme_pals2/project_dictate.aspx
18. Tagging Tool for EPrints repositories:
<http://www.connotea.org/taggingtool>
19. EPrints software:
<http://www.eprints.org/>
20. Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH):
<http://www.openarchives.org/>
21. OAI-compliant Metadata Repository for a Specialist Publisher of E-journals:
http://www.eevl.ac.uk/projects_503.htm
22. STARGATE (Static Repository Gateway and Toolkit). Project website:
<http://cdlr.strath.ac.uk/stargate/Index.htm>;
JISC web page:
http://www.jisc.ac.uk/whatwedo/programmes/programme_pals2/project_stargate.aspx
23. Specification for an OAI Static Repository and an OAI Static Repository Gateway:
<http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm>
24. TOCRoSS (Table of Contents by Really Simple Syndication):
http://www.jisc.ac.uk/whatwedo/programmes/programme_pals2/project_tocross.aspx
25. RSS 2.0 specification:
<http://www.rssboard.org/rss-specification>
26. ONIX for Serials SRN (Serials Release Notification):
http://www.editeur.org/onixserials/ONIX_SRN09.html
27. TOCRoSS Publisher Starter Kit:
<http://sourceforge.net/projects/tocross/>
28. SUNCAT:
<http://www.suncat.ac.uk/>
29. ONIX for Serials SOH (Serials Online Holdings):
http://www.editeur.org/onixserials/ONIX_SOH1.0.html
30. AIMSS (Automating Ingest of Metadata on Serials Subscriptions):
http://www.jisc.ac.uk/whatwedo/programmes/programme_pals2/project_aimss.aspx
31. Guy, F, Automating Metadata Loading for Serials Subscriptions: A Case Study Using ONIX for Serials

- (Serials Online Holdings) to Update Records on SUNCAT, *Serials*, 2006, 19(3), 220–229.
32. Low, B, Fergusson, D and MacColl, J, Resource discovery services for grid computing training, 2006. Paper presented at the International Conference on Web-Based Learning, July 2006. <http://baillie.lib.ed.ac.uk/metadataplus/icw12006b.pdf>.
33. metadata+ (Machine Services for Metadata Discovery and Aggregation). Project website: <http://baillie.lib.ed.ac.uk/metadataplus/> JISC web page: http://www.jisc.ac.uk/whatwedo/programmes/programme_pals2/project_metadata_plus.aspx
34. Discovery and Access: Standards and the Information Chain, London, 7 December 2006: <http://www.ukoln.ac.uk/events/pals-2006/>
35. Synthesis of PALS Metadata and Interoperability programmes: http://www.jisc.ac.uk/whatwedo/programmes/programme_pals2/synthesis

Article © Christine Baldwin

■ Christine Baldwin
Information Design & Management
Broom
Hinksey Hill
Oxford OX1 5BH, UK
E-mail: c.baldwin@dial.pipex.com

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

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=20&issue=1&spage=43>

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

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=20&issue=1>