Strathprints Institutional Repository

Robertson, R. John and , JISC (Joint Information Systems Committee) (Funder) (2006) *STARGATE: Static Repository Gateway and Toolkit. Final Project Report.* [Report]

JISC DEVELOPMENT PROGRAMMES

Project Document Cover Sheet

STARGATE FINAL REPORT

Project Acronym	STARGATE	**Project ID**	
Project Title	Static Repository Gateway and Toolkit: Enabling small publishers to participate in OAI-based services		
Start Date	1st Oct 2006	**End Date**	21 June 2006
Lead Institution	Centre for Digital Library Research, University of Strathclyde		
Project Director	Jane Barton/ Alan Dawson		
Project Manager & contact details	R. John Robertson Centre for Digital Library Research, Department of Computer & Information Sciences, University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow G1 1XH Tel: 0141 548 5854 Fax: 0141 548 4523 Email: robert.robertson@cis.strath.ac.uk		
Partner Institutions	Heriot Watt University *Journal of Digital Information* Professor Tom Wilson (*Information Research: an international electronic journal*) Library & Information Research Group (*Library & Information Research*) CILIPS/SLIC (*Information Scotland*)		
Project Web URL	http://cdlr.strath.ac.uk/stargate/		
Programme Name (and number)	JISC PALS II Metadata & Interoperability Projects		
Programme Manager	Christine Baldwin		

Document

Document Title	*Stargate Final Report*		
Author(s) & project role	R. John Robertson (Project Officer)		
Date	08/08/2006	**Filename**	StargateFinalReport1_4
URL	http://cdlr.strath.ac.uk/stargate/		
Access	□ Project and JISC internal	✓ General dissemination	

Document History

Version	Date	Comments
1.0	21/06/2006	Draft Final Report submitted to JISC and project partners
1.1	12/07/2006	Revised Final Report
1.2	19/07/2006	Appendix added
1.3	28/07/2006	Additional material and appendices added
1.4	08/08/2006	Minor revisions

Stargate Final Report

Author: R. John Robertson

Date: 8 August 2006
Version: 1.4
Produced as part of the STARGATE Project (http//:cdlr.strath.ac.uk/stargate/)

CENTRE FOR DIGITAL
LIBRARY RESEARCH
cdlr.strath.ac.uk

Table of Contents

Acknowledgements

This project was funded under the JISC PALS II Metadata & Interoperability Projects programme.

We would like to thank the journals and service who participated in the STARGATE project:

- *Information Research* published and edited by Professor Tom Wilson http://informationr.net/ir/.
- *Journal of Digital Information* (JoDI) published by Texas A&M University http://jodi.tamu.edu/.
- *Information Scotland* published by the Scottish Library and Information Council on behalf of the Chartered Institute of Library and Information Professionals (Scotland) http://www.slainte.org.uk/publications/serials/infoscot/contents.html.
- *Library and Information Research* (LIR) published by the Chartered Institute of Library and Information Professionals' Library and Information Research Group http://www.cilip.org.uk/specialinterestgroups/bysubject/research/publications/journal.
- TechXtra engineering information service based at Heriot Watt University http://www.techxtra.ac.uk/ .

We would also like to thank the others who contributed to the project, in particular the JISC programme manager and steering group, OAIster (a service based at the University of Michigan), and The Association of Learned and Professional Society Publishers (ALPSP).

Executive Summary

STARGATE (Static Repository Gateway and Toolkit) was funded by the Joint Information Systems Committee (JISC) and is intended to demonstrate the ease of use of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) Static Repository technology, and the potential benefits offered to publishers in making their metadata available in this way This technology offers a simpler method of participating in many information discovery services than creating fully-fledged OAI-compliant repositories. It does this by allowing the infrastructure and technical support required to participate in OAI-based services to be shifted from the data provider (the journal) to a third party and allows a single third party gateway provider to provide intermediation for many data providers (journals).

Specifically, STARGATE has created a series of Static Repositories of publisher metadata provided by a selection of Library and Information Science journals. It has demonstrated the interoperability of these repositories by exposing their metadata via a Static Repository Gateway for harvesting and cross-searching by external service providers. The project has conducted a critical evaluation of the Static Repository approach in conjunction with the participating publishers and service providers.

The technology works. The project has demonstrated that Static Repositories are easy to create and that the differences between fully-fledged and static OAI Repositories have no impact on the participation of small journal publishers in OAI-based services. The problems for a service that arise out of the use of Static Repositories are parallel to those created by any other repository dealing with journal articles. Problems arise from the diversity of metadata element sets provided by a given journal and the lack of specific metadata elements for the articles' volume and issue details. Another issue for the use of publishers' metadata arise as the collection policies of some existing services only allow Open Access materials to be included in them.

The project recommends that the use of Static Repositories continues to be explored – in particular as a flexible way to expose existing sets of structured information to OAI services and to create the opportunity to enhance the metadata as part of the process. The project further recommends that the publishing community consider the creation or adoption of an application profile for journal articles to support information discovery that can search by volume and issue. Significant further use of the Static Repository technology by small journal publishers will require the future creation and maintenance of a community-specific Static Repository Gateway. Further use will also require advocacy within the publishing community but might initially be most effectively kick-started through the creation of OAI repositories based on metadata held by the commercial services which publish or mediate access to electronic copies of journals on behalf of small publishers.

Background

The following background section assumes a degree of awareness with the concepts and technologies used in the Information Environment, these technologies and their origins are more fully explained in the introductory materials created by the project and appended at the end of this report. An article introducing the project and its relevance to small publishers is also appended.

The STARGATE project has explored the use of Static Repositories as a means of exposing publisher metadata to OAI-based disclosure, discovery and alerting services within the JISC Information Environment and beyond. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a domain-neutral technical solution for metadata harvesting provided by the Open Archives Initiative (OAI). OAI-PMH allows repositories to provide metadata about their content to external information-discovery services. The utility of OAI exposure for publishers' metadata was established by the PALS I project carried out at Heriot Watt University in 2003 [1].

Static Repositories were developed as an alternative to fully fledged OAI-compliant repositories [2] and have been implemented successfully within the Open Language Archives Community [3] as a means of widening participation in OAI-based systems and services.[1] Their use in the HaIRST project was highlighted as one of the key pieces of learning from the JISC's FAIR Programme [4]. Thus far their use within the Information Environment generally and the publishing community has been limited.

A normal OAI repository is a database whereas a static repository is a document. A database is more flexible and powerful than a document because you can submit queries to it and extract information from it in different ways, but many people find it easier to work with documents than databases. In particular, the databases used in OAI repositories, often based on Open Source software, require a degree of technical knowledge to install and maintain. A static repository is simply a document with a clear structure pre-defined for a specific purpose.

The static repository approach works by loading these structured documents (static repositories) into an external third-party database (a static repository gateway). It then becomes possible to extract information from the document via a normal OAI query. Static repositories have slight limitations on the functions they can support and have a limit on the number of records they can hold (~5000).

STARGATE has implemented a series of Static Repositories of publisher metadata for four journals, demonstrated the interoperability of the exposed metadata through harvesting and cross-searching via a Static Repository Gateway, and conducted a critical evaluation of the approach with publishers and service providers.

[1] It should be noted the phrase OAI repository refers exclusively to normal repositories (i.e. fully functional database-driven).

The project brought together an experienced technical team and a group of publishers representative of the type that, it was believed, might require a simpler approach to the exposure of its metadata than is offered by the full OAI repository, and yet would be sufficiently familiar with the concepts of metadata-driven information access, standards compliance and interoperability to be able to contribute to an informed evaluation of the approach adopted within the project.

Aims and Objectives

Aim

The project's primary aim was to lower the technical barriers to the implementation of OAI-compliant repositories, thereby enabling small publishers of electronic resources to participate more readily in OAI-based disclosure and delivery services within the JISC Information Environment and beyond.

Objectives

1. To build an interoperability demonstrator based around the Static Repository Gateway developed in the HaIRST project and dovetailing with the OAI-based architecture developed in PALS I.

   It will incorporate a series of Static Repositories containing metadata from a range of small e-journal publishers, and will enable an exploration of the issues surrounding the deployment of this technology and its applicability to the exposure of publisher metadata via OAI-PMH.

   It will also expose publisher metadata via the Static Repository Gateway to the HaIRST ARC harvester [5] for aggregation with other publisher metadata and other related repositories, and to the EEVL Xtra service [6] [*this became the TechXtra service [7] over the course of the project*] for cross-searching alongside a range of relevant resources. Other opportunities to test and/or demonstrate the interoperability of the exposed metadata will be sought during the project. [*The Static Repositories were also harvested and cross-searched by OAIster [8] – a general OAI-based service – and will also be harvested by Metalis [9] – a subject specific OAI-based service*]

   Static Repositories will initially be hosted and managed centrally, although by the end of the project it is likely that at least one will be hosted and managed by the publisher [*Information Scotland has taken over the generation of its Static Repository, but has not yet updated it*]. Persistent URIs will be used to provide access to the full text of articles, which will continue to be hosted by the publisher, although if possible the potential for the full text to be incorporated into the Static Repository will be explored [*Mid-project it was decided that the project should not attempt to assign persistent URI's as it could not sustain them; exploring full text incorporation was not possible. Stargate was, however, able to significantly extend the scope of the project and create Static Repositories for the entire run of three of the participating journals*].

2. To develop tools and guidelines that will facilitate the implementation of Static Repositories by small publishers that wish to expose their metadata to OAI-based services.

   Tools to be developed within the project will include documented case studies as described above, scripts to convert various metadata sources into OAI-compliant Static Repositories, and templates to create DC-compliant metadata sources suitable for conversion [*sources of structured information are more usefully turned into DC compliant metadata via a database and script as this allows better metadata manipulation*].

   Guidelines will take the form of a critical analysis of the Static Repositories approach to exposing publisher metadata, including an evaluation of project experiences, a comparison with alternative approaches to exposing publisher metadata, and recommendations on how and in what circumstances publishers and/or services might implement the Static Repositories approach [In additional to anticipated publisher input into these guidelines, the project benefited from the input of ALPSP on these issues].

These tools and guidelines would not only form the basis of a toolkit for publishers, but would also be a valuable resource for services that seek to broaden the range of information available to their users, either by advocacy or by providing technical support and/or hosting facilities for publishers.

Methodology

The overall approach of the project was to use existing structured information from a publisher and to help create a Static Repository. This was implemented by using a simple database to store and manipulate the publisher's information. The database also stored the information necessary to generate the Static Repositories.

There are at least three ways to generate a Static Repository (in essence an XML file); it could be hand-coded, generated from an existing XML file by using an XSLT (a language for transforming one XML file into another), or generated from a database of some kind. Hand coding is not a viable option as it would require an unreasonable amount of data entry and a significant understanding of the technologies involved. XSLT is a reasonable option and could be used to create a Static Repository. It would, however, require a degree of hand-coded customisation for each new journal, and it would require the information used to create the Static Repository to have an entirely consistent structure and be in well-formed XML. In contrast, using a database allows the structured information to require a degree of adjustment and manipulation (such as the addition of data) and to be provided in a variety of forms (not just XML). It also permits the system to be adapted for a new journal by editing database fields rather than adding code or mark-up.

The initial phase of the project used metadata supplied by the participating journals or captured it from their live websites. An Access module (a Visual Basic programme) was written to process this metadata into an Access database (containing appropriate tables), the structure of the OAI protocol was also created in the database, and a second module was written to extract all of this information and create a Static Repository for each of the journals. It quickly became apparent that a degree of data manipulation would be required – eventually this was implemented by developing profiles for each journal which contain metadata common to each of their articles and also creating a table of transformations that need to be carried out on a given data set. These were built into the database so that changing, or adding to, the values in them could be carried out without having to change the modules. Although adding to these tables requires a degree of analysis, the process does not require technical knowledge and so it should be accessible to most non-technical users. This is similar to the reason that proprietary desktop software (i.e. MS Access) was used – such software supports novice users, is capable of carrying out the desired tasks, and is ubiquitous on business computers.

These Static Repositories were then registered with a Gateway to allow them to be queried and harvested by the OAI-PMH protocol. They were then harvested by the HaIRST harvester, the TechXtra service, the OAIster service, and made available to the wider Information Environment.

The project's technical development was paralleled by a series of meetings with the participating journals. These meetings allowed a discussion of the workflows and standards used by the journals to create the structured information, and supported reflection on the process of setting up the repositories, the appropriateness of the technology for small publishers, and the value of publisher participation in OAI-based services. Dissemination of information about the project also elicited feedback from the publishing and technical communities about a number of these issues.

The methodology proved effective at creating and assessing the utility of Static Repositories. It had the incidental additional benefit of supporting a degree of metadata cleaning and manipulation. The ability to gain an overview of the metadata from the participating journals also allowed some consideration of how the data sets might interoperate in a service. The methodology required a degree of adjustment, however, in so far as a number of the publishers could not attend a physical meeting so the project team held online meetings with them. The project was able to demonstrate the effectiveness of utilising Static Repositories in terms of ease of use, but an assessment of the effectiveness in terms of dissemination in the wider Information Environment will only be provable over time as such repositories are in use

Implementation

The project began by gathering metadata from the participating publishers. Three of the publishers sent their metadata, and the metadata for *Information Research* was downloaded. These metadata sets were reviewed to identify the consistent record and field delimiters. Although it would be possible to review the metadata thoroughly record by record, one of the key issues at this stage of the project was to carry out a 'disposable' data import and get the data into a table so that an overview of it could be obtained, as this would allow us to see which metadata elements where consistently present in each of the journals and also let us note any metadata that was identical across an entire journal issue. The disposable approach allows a more informed database and module design.

Based on the information gathered from this overview the project staff then created a database with a number of tables and wrote a module to read the structured data into the tables. At this stage of the development the module had customised journal-specific instructions. The approach taken by the project to metadata import and manipulation is that, where possible, any changes that are made to metadata are repeatable so that where possible the process can always start from the original structured information (later in the project when changes to the original information were appropriately required these changes were passed onto the publishers so that they could be reflected in the source data).

The import module was refined and a table containing the structure of an OAI Static Repository and a macro to create it were imported from earlier development work in the HaIRST project. This was then used to create initial OAI Static Repository files. These initial files had some difficulties with characters from the metadata values that were illegal in XML. This problem was addressed by adding mappings to equivalent valid character values to the relevant subroutines in the import module . These tweaks allowed the creation of valid Static Repositories for the journals. The repositories were then registered in a local demonstration Static Repository Gateway and harvested by a local harvester. During this stage of the project it became clear that there was only a small set of metadata elements that was common to all the participating journals and that despite being from the same subject area they had a made significantly different choices about the metadata elements they opted to include. This diversity would to some degree limit the search refinements that a service could implement. Of particular importance to a service was that the URL of the paper was often missing from the metadata. The project also created custom components for each journal to extract the volume and issue details for each article into their own database fields. Although these are not required within Dublin Core, project staff felt that they would be key components of an article specific service and demonstrated their extraction (typically from the paper's URL); the volume and issue details were inserted into the DC relation field for exposure – this would allow their use by a service as part of a general keyword search.

Both sets of meetings with the journal publishers had been planned as physical meetings, but for a variety of reasons ended up being a series of online meetings. This alternative was able to ensure that the meetings continued as planned but had the side-effect of increasing the amount of time required for them. It also allowed a more technically interactive discussion as both parties could see technical issues illustrated in real time and interact with relevant online resources. Due to unforeseen circumstances, one of the journals, Library and Information Research, was unable to participate in the project beyond supplying its initial metadata. This reduced the level of input the project gained from the publishing community. During the meetings it emerged that the publishers of Information Scotland were intending to continue to use the developed tools after the project to produce Static Repositories. In response to this the project produced a slightly more detailed case-study of the production of their repository to support such use.

The writing and disseminating aspects of the project provided unexpected benefits – a number of interested observers contacted the project about its work. In particular, messages to a number of email lists produced two very significant contacts for the project. The first of these was from the OAIster service and expressed interest in incorporating the Static Repositories into their service. The second was from the Association of Learned and Professional Society Publishers (ALPSP). The contact with ALPSP was followed up by a conference call which allowed a discussion of the utility of Static Repositories and outlined issues that might arise in communicating this to the publishing community. The involvement of ALPSP significantly improved the project’s overview of the publishing community.

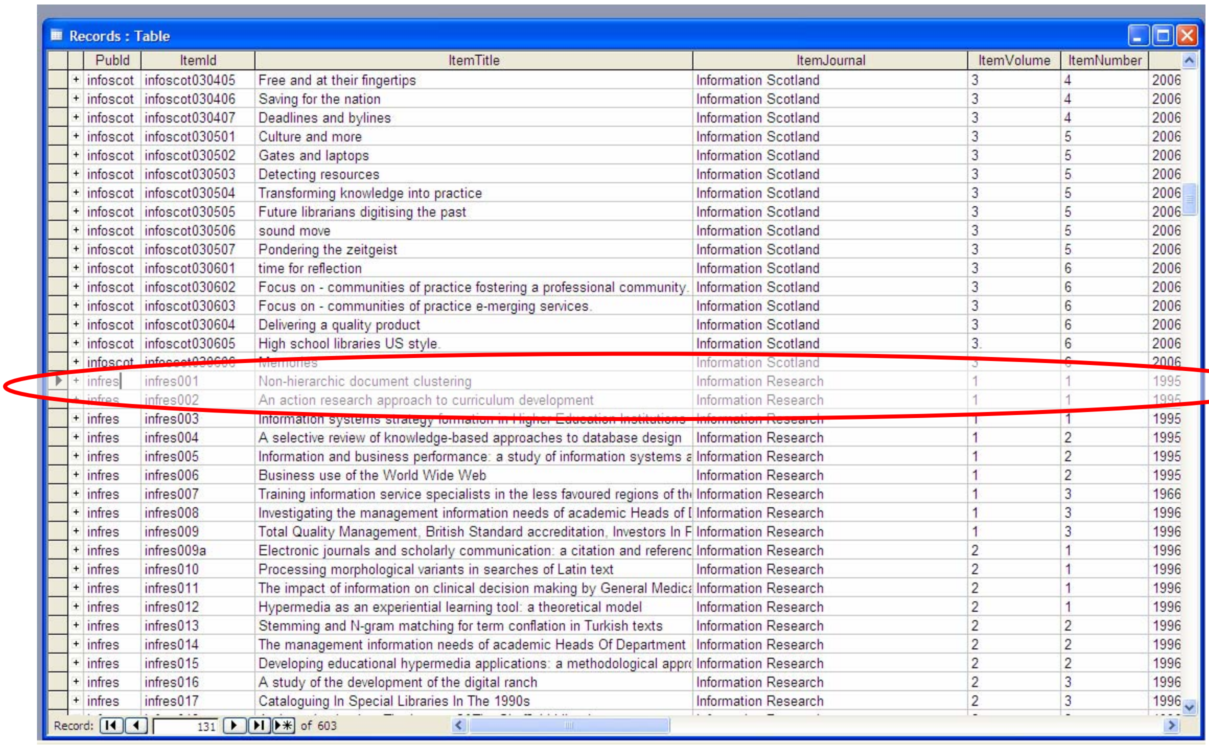
As a result of these meetings, especially the feedback from the services TechXtra and OAIster, STARGATE began, with the journals’ permission, to produce Static Repositories for the entire run of *Information Research*, *Information Scotland*, and the *Journal of Digital Information*. An initial review of the metadata involved led to some refinement of the database design – notably, the translations required to map invalid characters to their XML values was removed from the module and instead put into a table of character translations that the module refers to. This produced a much more elegant solution and allows less technical users to add to a table if the journal requires a translation for a previously unused character, rather than having to change or add to code. The data mappings were also moved from the module into this table. This allowed the easier creation of new character mappings if a new journal were to be added.

With the creation of full repositories the number of data errors to be addressed increased. Most of these arose from changes in metadata practice as the journals matured. Some of these required a degree of metadata editing. The project was able to assist publishers carry out some of this editing of errors in the original sources. In cases where practice had changed over time, however, the project had to implement changes on local data, as the cost of adjusting legacy data was outwith the scope of publisher involvement. This was carried out by making changes to the concatenated metadata files as, once identified, this allowed structural issues to be changed *en masse*. The flexibility of using a database to process the metadata allowed errors and different practices to be easily identified.

The repositories are now available via the Stargate website and are searchable in the TechXtra service (until the end of July 2006) and the OAIster service.

Outputs and Results

The project has produced four Static Repositories from publishers' metadata. Three of these contain metadata for the entire run of their respective journal. These repositories are registered in a Gateway and exposed for harvest and have been successfully incorporated into two services.



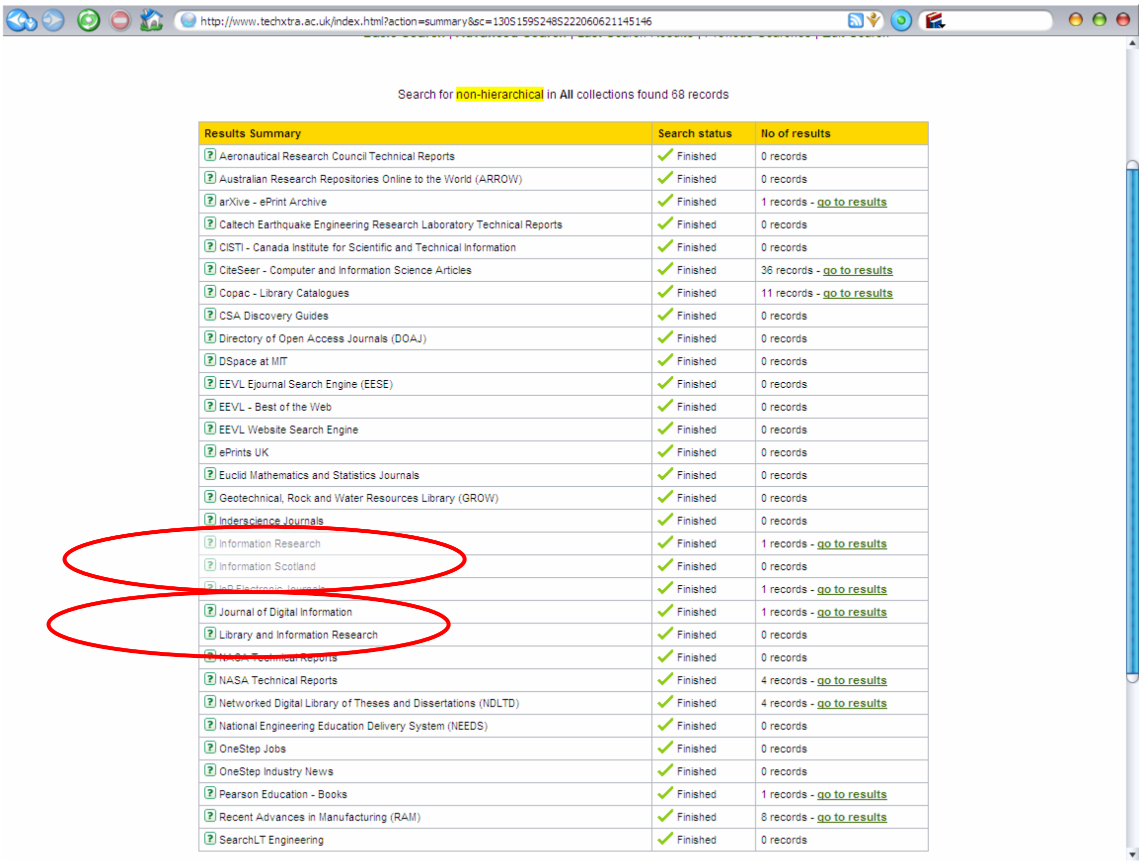
PubId	ItemId	ItemTitle	ItemJournal	ItemVolume	ItemNumber	
infoscot	infoscot030405	Free and at their fingertips	Information Scotland	3	4	2006
infoscot	infoscot030406	Saving for the nation	Information Scotland	3	4	2006
infoscot	infoscot030407	Deadlines and bylines	Information Scotland	3	4	2006
infoscot	infoscot030501	Culture and more	Information Scotland	3	5	2006
infoscot	infoscot030502	Gates and laptops	Information Scotland	3	5	2006
infoscot	infoscot030503	Detecting resources	Information Scotland	3	5	2006
infoscot	infoscot030504	Transforming knowledge into practice	Information Scotland	3	5	2006
infoscot	infoscot030505	Future librarians digitising the past	Information Scotland	3	5	2006
infoscot	infoscot030506	sound move	Information Scotland	3	5	2006
infoscot	infoscot030507	Pondering the zeitgeist	Information Scotland	3	5	2006
infoscot	infoscot030601	time for reflection	Information Scotland	3	6	2006
infoscot	infoscot030602	Focus on - communities of practice fostering a professional community.	Information Scotland	3	6	2006
infoscot	infoscot030603	Focus on - communities of practice e-merging services.	Information Scotland	3	6	2006
infoscot	infoscot030604	Delivering a quality product	Information Scotland	3	6	2006
infoscot	infoscot030605	High school libraries US style.	Information Scotland	3.	6	2006
infoscot	infoscot030606	Memories	Information Scotland	3	6	2006
infres	infres001	Non-hierarchic document clustering	Information Research	1	1	1995
infres	infres002	An action research approach to curriculum development	Information Research	1	1	1995
infres	infres003	Information systems strategy formation in Higher Education Institutions	Information Research	1	1	1995
infres	infres004	A selective review of knowledge-based approaches to database design	Information Research	1	2	1995
infres	infres005	Information and business performance: a study of information systems a	Information Research	1	2	1995
infres	infres006	Business use of the World Wide Web	Information Research	1	2	1995
infres	infres007	Training information service specialists in the less favoured regions of th	Information Research	1	3	1966
infres	infres008	Investigating the management information needs of academic Heads of [	Information Research	1	3	1996
infres	infres009	Total Quality Management, British Standard accreditation, Investors In F	Information Research	1	3	1996
infres	infres009a	Electronic journals and scholarly communication: a citation and referenc	Information Research	2	1	1996
infres	infres010	Processing morphological variants in searches of Latin text	Information Research	2	1	1996
infres	infres011	The impact of information on clinical decision making by General Medic	Information Research	2	1	1996
infres	infres012	Hypermedia as an experiential learning tool: a theoretical model	Information Research	2	1	1996
infres	infres013	Stemming and N-gram matching for term conflation in Turkish texts	Information Research	2	2	1996
infres	infres014	The management information needs of academic Heads Of Department	Information Research	2	2	1996
infres	infres015	Developing educational hypermedia applications: a methodological appr	Information Research	2	2	1996
infres	infres016	A study of the development of the digital ranch	Information Research	2	3	1996
infres	infres017	Cataloguing In Special Libraries In The 1990s	Information Research	2	3	1996

A sample record within the database

```
<oai:record>
<oai:header>
<oai:identifier>oai:infres001</oai:identifier>
<oai:datestamp>2006-05-30</oai:datestamp>
</oai:header>
<oai:metadata>
<oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/
http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:title>Non-hierarchic document clustering</dc:title>
<dc:creator>
Gareth Jones, Alexander M. Robertson, Chawchat Santimetvirul, Peter Willett
</dc:creator>
<dc:publisher>Professor T.D. Wilson, University of Sheffield</dc:publisher>
<dc:date>1995</dc:date>
<dc:description>
Cluster analysis, or automatic classification, is a multivariate
statistical technique that seeks to identify groups, or clusters, of
similar objects in a multi-dimensional space. There have been many attempts
over the years to use such procedures for the organisation of document
databases, so that documents with large numbers of index terms in common
are grouped together. In this paper, we consider the use of a genetic
algorithm, henceforth a GA, for document clustering. GAs are a class of
non-deterministic algorithms that derive from Darwinian theories of
evolution. They provide good, though not necessarily optimal solutions to
combinatorial optimisation problems, where the number of possible solutions
```

```
is far too great for all of the possibilities to be explored in a
reasonable time by a deterministic algorithm. One such problem is that of
non-hierarchic clustering, where the clustering method seeks to partition a
set of objects into a set of non-overlapping groups so as to maximise some
external criterion of 'goodness of clustering', typically the extent to
which the within-cluster inter-object similarities are maximised and the
between-cluster similarities minimised.
</dc:description>
<dc:identifier>http://informationr.net/ir/1-1/paper1.html</dc:identifier>
<dc:relation>
Published in Information Research Volume 1 Number 1
</dc:relation>
<dc:subject>
information retrieval, document clustering, clustering, algorithms, cluster
analysis, automatic classification, genetic algorithms
</dc:subject>
</oai_dc:dc>
</oai:metadata>
</oai:record>
```

Sample of a Static Repository record

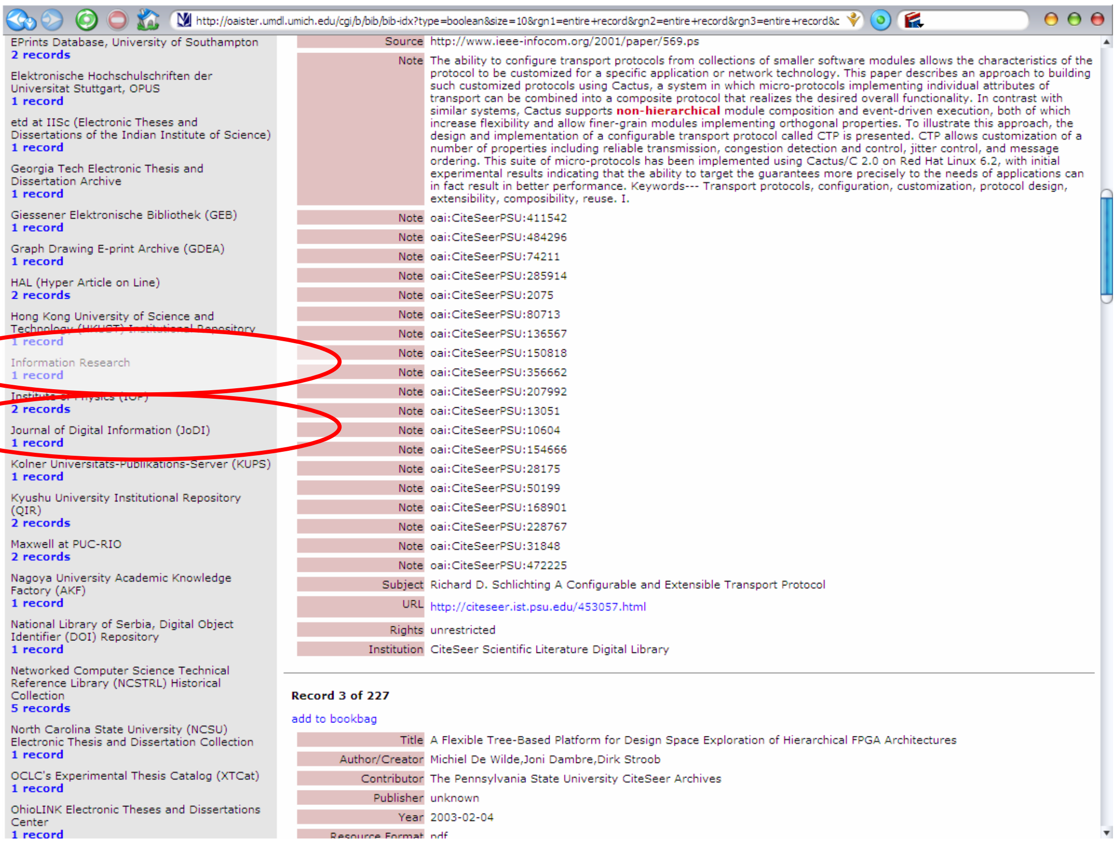


http://www.techxtra.ac.uk/index.html?action=summary&sc=130S159S248S222060621145146

Search for non-hierarchical in **All** collections found 68 records

Results Summary	Search status	No of results
Aeronautical Research Council Technical Reports	✓ Finished	0 records
Australian Research Repositories Online to the World (ARROW)	✓ Finished	0 records
arXive - ePrint Archive	✓ Finished	1 records - go to results
Caltech Earthquake Engineering Research Laboratory Technical Reports	✓ Finished	0 records
CISTI - Canada Institute for Scientific and Technical Information	✓ Finished	0 records
CiteSeer - Computer and Information Science Articles	✓ Finished	36 records - go to results
Copac - Library Catalogues	✓ Finished	11 records - go to results
CSA Discovery Guides	✓ Finished	0 records
Directory of Open Access Journals (DOAJ)	✓ Finished	0 records
DSpace at MIT	✓ Finished	0 records
EEVL Ejournal Search Engine (EESE)	✓ Finished	0 records
EEVL - Best of the Web	✓ Finished	0 records
EEVL Website Search Engine	✓ Finished	0 records
ePrints UK	✓ Finished	0 records
Euclid Mathematics and Statistics Journals	✓ Finished	0 records
Geotechnical, Rock and Water Resources Library (GROW)	✓ Finished	0 records
Inderscience Journals	✓ Finished	0 records
Information Research	✓ Finished	1 records - go to results
Information Scotland	✓ Finished	0 records
IoP Electronic Journals	✓ Finished	1 records - go to results
Journal of Digital Information	✓ Finished	1 records - go to results
Library and Information Research	✓ Finished	0 records
NASA Technical Reports	✓ Finished	0 records
NASA Technical Reports	✓ Finished	4 records - go to results
Networked Digital Library of Theses and Dissertations (NDLTD)	✓ Finished	4 records - go to results
National Engineering Education Delivery System (NEEDS)	✓ Finished	0 records
OneStep Jobs	✓ Finished	0 records
OneStep Industry News	✓ Finished	0 records
Pearson Education - Books	✓ Finished	1 records - go to results
Recent Advances in Manufacturing (RAM)	✓ Finished	8 records - go to results
SearchLT Engineering	✓ Finished	0 records

Screenshot of Static Repositories in the TechXtra service



http://oaister.umdl.umich.edu/cgi/b/bib/bib-idx?type=boolean&size=10&rgn1=entire+record&rgn2=entire+record&rgn3=entire+record&c

EPrints Database, University of Southampton
2 records

Elektronische Hochschulschriften der Universitat Stuttgart, OPUS
1 record

etd at IISc (Electronic Theses and Dissertations of the Indian Institute of Science)
1 record

Georgia Tech Electronic Thesis and Dissertation Archive
1 record

Giessener Elektronische Bibliothek (GEB)
1 record

Graph Drawing E-print Archive (GDEA)
1 record

HAL (Hyper Article on Line)
2 records

Hong Kong University of Science and Technology (HKUST) Institutional Repository
1 record

Information Research
1 record

Institute of Physics (IOP)
2 records

Journal of Digital Information (JoDI)
1 record

Kolner Universitats-Publikations-Server (KUPS)
1 record

Kyushu University Institutional Repository (QIR)
2 records

Maxwell at PUC-RIO
2 records

Nagoya University Academic Knowledge Factory (AKF)
1 record

National Library of Serbia, Digital Object Identifier (DOI) Repository
1 record

Networked Computer Science Technical Reference Library (NCSTRL) Historical Collection
5 records

North Carolina State University (NCSU) Electronic Thesis and Dissertation Collection
1 record

OCLC's Experimental Thesis Catalog (XTCat)
1 record

OhioLINK Electronic Theses and Dissertations Center
1 record

Source	http://www.ieee-infocom.org/2001/paper/569.ps
Note	The ability to configure transport protocols from collections of smaller software modules allows the characteristics of the protocol to be customized for a specific application or network technology. This paper describes an approach to building such customized protocols using Cactus, a system in which micro-protocols implementing individual attributes of transport can be combined into a composite protocol that realizes the desired overall functionality. In contrast with similar systems, Cactus supports **non-hierarchical** module composition and event-driven execution, both of which increase flexibility and allow finer-grain modules implementing orthogonal properties. To illustrate this approach, the design and implementation of a configurable transport protocol called CTP is presented. CTP allows customization of a number of properties including reliable transmission, congestion detection and control, jitter control, and message ordering. This suite of micro-protocols has been implemented using Cactus/C 2.0 on Red Hat Linux 6.2, with initial experimental results indicating that the ability to target the guarantees more precisely to the needs of applications can in fact result in better performance. Keywords--- Transport protocols, configuration, customization, protocol design, extensibility, composibility, reuse. I.
Note	oai:CiteSeerPSU:411542
Note	oai:CiteSeerPSU:484296
Note	oai:CiteSeerPSU:74211
Note	oai:CiteSeerPSU:285914
Note	oai:CiteSeerPSU:2075
Note	oai:CiteSeerPSU:80713
Note	oai:CiteSeerPSU:136567
Note	oai:CiteSeerPSU:150818
Note	oai:CiteSeerPSU:356662
Note	oai:CiteSeerPSU:207992
Note	oai:CiteSeerPSU:13051
Note	oai:CiteSeerPSU:10604
Note	oai:CiteSeerPSU:154666
Note	oai:CiteSeerPSU:28175
Note	oai:CiteSeerPSU:50199
Note	oai:CiteSeerPSU:168901
Note	oai:CiteSeerPSU:228767
Note	oai:CiteSeerPSU:31848
Note	oai:CiteSeerPSU:472225
Subject	Richard D. Schlichting A Configurable and Extensible Transport Protocol
URL	http://citeseer.ist.psu.edu/453057.html
Rights	unrestricted
Institution	CiteSeer Scientific Literature Digital Library

Record 3 of 227

add to bookbag

Title	A Flexible Tree-Based Platform for Design Space Exploration of Hierarchical FPGA Architectures
Author/Creator	Michiel De Wilde,Joni Dambre,Dirk Stroob
Contributor	The Pennsylvania State University CiteSeer Archives
Publisher	unknown
Year	2003-02-04
Resource Format	pdf

Screenshot of Static Repositories in the OAIster service

In the course of producing these repositories the project has demonstrated the utility of Static Repository technology as a low-cost option for participation in OAI-based services. It has imported and processed metadata in a variety of formats and from a variety of sources. In doing so, the project has been able to demonstrate and investigate something about the nature of managing metadata and how a service can supplement existing metadata.

The project has produced initial tools and case studies of the process to support the use of this technology by others. Part 1 and 3 of the tools are appended to this report and all the tools (and accompanying database) are available on the project website (http://cdlr.strath.ac.uk/stargate/tools.htm). The tools are:

1) **Stargate Tools Part 1: Introduction, key issues and relevant technologies** (version 1.4)
   This document briefly introduces the project and technologies used to improve access to information about articles. It introduces the Open Archives Initiative Protocol for Metadata Harvesting and the Dublin Core Metadata Element Set.
2) **Stargate Tools Part 2: Creating Static Repositories, evaluating this method, and case studies** (version 1.3)
   This document outlines and details a method of producing static repositories from existing structured information (e.g. <meta> tags in a web page). It also provides advice for metadata normalisation, an evaluation of this method of participation in OAI services, and four case studies of producing static repositories for the participating journals.

3) **The database tool and VB modules**
   This is the Access database and modules used to create static repositories. Its use is detailed in the above guide: Stargate Tools Part 2.

4) **Stargate Tools Part 3: Issues for developing the use of OAI-based services by the publishing community** (version 1.3)
   This document briefly outlines the issues the project encountered in creating static repositories that would impact on widespread use of this technology by publishers. It also describes the conditions under which such static repositories could be of use and what benefit might be derived from publisher participation in OAI-based services. Many of these issues apply to any community considering the use of static repositories.

Outcomes

Achieving the aim

The project's primary aim was to lower the technical barriers to the implementation of OAI-compliant repositories, thereby enabling small publishers of electronic resources to participate more readily in OAI-based disclosure and delivery services within the JISC Information Environment and beyond. It has provided a proof of concept example for the use of this technology and sample tools to assist others to more easily participate in OAI- based services.

Meeting objectives

1. The project successfully built an interoperability demonstrator extending and combining some of the work of previous projects (the HaIRST project and OAI-compliant Metadata Repository for a Specialist Publisher of E-journals project).

   It produced a series of Static Repositories containing metadata from a range of small e-journal publishers, and explored issues surrounding the deployment of this technology and its applicability to the exposure of publisher metadata via OAI-PMH.

   It carried out demonstration harvesting and cross-searching via the HaIRST ARC harvester and the TechXtra service, and was also harvested by the OAIster service.

   Post-project the Information Scotland Static Repository will be hosted and managed by the publisher and the Information Research Static Repository may continue be maintained. It was decided that the project should not attempt to assign persistent URI's as it could not sustain them. Stargate was, however, able to significantly extend the scope of the project and create Static Repositories for the entire run of three of the participating journals.

2. The project developed tools and guidelines to facilitate the implementation of Static Repositories by small publishers that wish to expose their metadata to OAI-based services.

   It developed documented case studies of the production of each Static Repository, scripts and a database to convert various metadata sources into OAI-compliant Static Repositories, and a technique to create DC-compliant metadata from structured information sources.

   It critically analysed the Static Repositories approach to exposing publisher metadata by engaging with the publishing community to produce guidelines about the possible use of the Static Repositories, evaluating project experiences, and alternative approaches.

The project has demonstrated the simplicity of participation in OAI-based services by using the Static Repository technology. The repositories developed by the project serve a useful purpose and may improve access to these journals, but the wider impact of the project is dependent on the degree to which the publishing community is aware of and willing to adopt such technologies. The project offers the basis of a desktop solution to a technical problem. For publishers without extensive technical support (or those with existing systems which do not offer OAI exposure), participation in OAI-based services is now a possibility.

If the technology is taken up by the publishing community, the teaching, learning, or research communities would benefit from the ability of service developers and providers to incorporate information about a wider range of journal articles than are currently available to them through subject-based or institutional repositories. Currently users choose to search for a particular publisher's content on the publisher's own system or to search for Open-Access content through a general service provider. Publisher participation would enable information about all of this content to be found via a single interface.

The project has further implications for other parts of the teaching, learning, or research communities in so far as it demonstrates the ability of Static Repositories to easily expose metadata about a set of resources. This technology lends itself to other applications such as the exposure of metadata about virtual museum objects or learning objects.

The chosen method of using an Access database proved very suitable to the task. More generally, the method worked well but required two significant changes. The more significant of these was that it was unrealistic to expect operational services to harvest and use the single-issue sample data. In response to feedback, the project had to prioritise the production of full Static Repositories for some of the journals. This also provided a functional addition to the Information Environment. A related issue is that the project underestimated how long it is likely to take to demonstrate the benefits of exposure of metadata to services – we anticipate that making this case will require a brief return to two of the journals to get access statistics in 6 months at the least. The less significant change was that we had to change the manner in which publishers participated in discussions with the project. This proved helpful for discussing technical aspects of the project but made discussing some of the community dimensions more difficult than anticipated.

Conclusions

The project proved the flexibility of Static Repositories to provide a low-cost entry to the Information Environment for small publishers. It also suggests the suitability of this technology for the exposure information about other assets

The creation of the repositories for entire journal runs helps preserve their identity within aggregated services and allows the development of themes within a journal to be tracked to a degree. Services based specifically on publishers' metadata should consider the development of an application profile that would allow volume and issue metadata to be included and used as search variables. Such an application profile might be similar to the ePrints application profile that is currently under development, and might support close interoperability with it. Interoperability between such profiles would allow refined cross-searching for journal articles from both publisher and ePrints communities. Although the majority of project participants feel strongly that the availability of publisher metadata would be beneficial to the Information Environment, a related issue is that many of the existing OAI-services only accept metadata from Open Access sources – with obvious implications for the dissemination of commercial publisher metadata. The participating publishers all offer, at least some, Open Access content and so could participate, but this remains a question for future service developers.

The project also notes that metadata variability will continue to be an issue but that services can, with a degree of effort, enhance metadata at the point of harvest to improve how metadata from different sources interoperates. Even within the four journals participating in the project there was significant variety in the metadata elements recorded, and within a given journal there was often variation over time in what metadata was available and how it was recorded.

Implications

Static Repositories should be more visible as a technology, especially to content development programmes and those using existing systems which do not have built-in OAI exposure. Deploying this type of database driven Static Repository generation has a lot of potential to integrate with existing systems and with established workflows. As a technology it also has the potential to bring OAI exposure to structured data held in legacy systems such as those built-on Access databases. It is suggested that the Static Repository approach represents a cost effective method of exposing this data.

Although STARGATE has demonstrated the use of Static Repositories as a low barrier method of participation they also have a potential function as a component in a larger system which adds OAI exposure as a short-term solution, pending the possible development of an integrated OAI export function – to an extent the project has done this in the case of Information Scotland for the Connexion system.

In addressing publishers' metadata specifically, one avenue for a future development of the service would be to examine other ways of getting access to publishers' metadata. This project's audience were small journal publisher who to, some degree, self-published. The project was reminded that a number of small journals contract out their publication process and that another potentially effective approach to exposing their metadata would be to collaborate with such commercial publishing services. This might allow the rapid production of a seed collection of publisher metadata for services.

The possible future development of a publisher-specific OAI application profile and service similar to that developed within the OLAC community, would allow the development of specialist information discovery services. Such customised services would benefit from the development of an extended form of oai_dc – for example, one that distinguished between date types and explicitly supported volume and issue elements. Such an application profile is likely to be similar to the one under development by the ePrints community and it is assumed that metadata in either application profile would be, on the whole, interoperable and permit the development of specialist information discovery services built on the metadata of both communities. One precondition of the further development of this technology within the publishing community would be the creation and ongoing support of a Static Repository Gateway for publisher's Static Repositories - they are currently in a research/ demonstration Gateway and other existing Gateways hold a great diversity of materials.

In the longer–term the number of services using OAI-PMH is likely to increase; it is becoming an essential part of the information environment and many new technical developments intend to offer the option to expose metadata via it. Of particular relevance to the publishing community is that many ongoing open-source journal workflow systems are beginning to incorporate it. In the future as such workflow systems are likely to move towards being offered as externally hosted web-based system, their ease of use and other features may prove an attractive option for even small publishers. They will, however, have difficulties with legacy data and it is anticipated that in this area, among others, Static Repositories will have an ongoing role. In particular, approaches like the one taken by this project offer the possibility of some degree of retrospective metadata manipulation as part of the repository generation process. It is felt that metadata manipulation is likely to remain fundamental to the incorporation of repositories in communal services.

Recommendations

1) To support the continued exploration and wider use of static repositories in the Information Environment (IE) by the wider community as well as the publishing community JISC should fund a static repository gateway with some degree of service guarantee. The gateways which exist are run by researchers or developers and are provided as 'experimental' – they make no guarantee about the service's provision or reliability. They are, therefore, not conducive to supporting service development or operation. It is recommended that a gateway be provided as part of the IE infrastructure (for example by, Edina or Mimas).

2) The publishing community, in so far as it has a communal interest in OAI exposure, should consider funding a static repository gateway with some degree of service guarantee. It would support the development of publisher-specific services and allow a degree of branding to be attached to the metadata provided therein. At this stage, it is unclear who would host or support this but there are a number of companies who already provide technical services to the sector (the gateway software required is open source, stable, and complete).

3) There is a degree of confusion within the publishing community about OAI. Those interested in supporting the use of this technology (both normal and static repositories) should fund the development of tailored materials and events to support dissemination and awareness-raising. It is suggested that JISC should consider participating in or providing such funding because of the benefits for service development and implementation that the exposure of publisher metadata would bring to the IE. It is, however, recommended that any such advocacy be carried out in partnership with, or led by, representatives of the publishing community.

   Such advocacy would benefit from further case studies of the use of OAI technologies by publishers. Although it is not JISC's business to fund the development of publisher's repositories, some form of collaboration would enhance the available evidence supporting the case for OAI exposure and any metadata exposed as a result would benefit the JISC community).

4) Many small- to medium-sized publishers use third party services to provide electronic access to their articles. It is suspected that such service providers store structured information for such articles which could be readily and rapidly exposed via OAI-PMH. In collaboration with the publishing community this possibility should be explored. Such an exploration should note any effect it might have on the business models of third-party service providers or abstracting and indexing services.

5) Metadata manipulation and harmonization at the service-provider level will continue to be essential to resource discovery in the IE. Taking the project's experience as a sample, the significant variation of metadata in use from even a relatively narrow spectrum of journals displays, JISC should continue to fund research and development that will support services as they try to cope with metadata from different sources.

References

[1] PALS I Project: OAI-compliant Metadata Repository for a Specialist Publisher of E-journals. http://www.eevl.ac.uk/projects_503.htm
[2] Hochstenbach, P. et al. (2003) *The OAI-PMH static repository and static repository gateway*. Available from: http://public.lanl.gov/herbertv/papers/jcdl2003-submitted-draft.pdf
[3] Open Language Archives Community (OLAC) Static Repository Gateway. http://www.language-archives.org/sr
[4] Brophy, P. (2005) *HaIRST Project summative evaluation: report*. Manchester: CERLIM.
[5] HaIRST arc harvester. http://speirserver.cdlr.strath.ac.uk:8088/arc/hairst_search.jsp
[6] MacLeod, Roddy. (2005) *EEVL* X*tra: the hidden web at your fingertips*. Ariadne, Issue 44, July 2005. Available from: http://www.ariadne.ac.uk/issue44/eevl/.
[7] TechXtra http://www.techxtra.ac.uk/
[8] OAIster http://oaister.umdl.umich.edu/o/oaister/
[9] Metalis http://metalis.cilea.it/

Appendix 1

Stargate: Exploring Static Repositories for Small Publishers

R. John Robertson introduces a project examining the potential benefits of OAI-PMH Static Repositories as a means of enabling small publishers to participate more fully in the information environment.[2]

Introduction

With the wider deployment of repositories, the Open Archives Initiative - Protocol for Metadata Harvesting (OAI-PMH) is becoming a common method of supporting interoperability between repositories and services. It provides 'an application-independent interoperability framework based on *metadata harvesting*' [1]. Nodes in a network using this protocol are 'data providers' or 'service providers'.

Although repository software supporting OAI-PMH is not overly complex [2], without programming skills or access to technical support, implementing and supporting a repository is not an entirely straightforward task. Static repositories and static repository gateways [3] are a development of the OAI-PMH specification that makes participation in networks of data and service providers even simpler. In essence a static repository is an XML file publicly available online at a persistent address. This file is registered in a static repository gateway which then presents it as a (slightly limited) OAI-PMH data provider.

One community that the static repository approach might benefit is the community of small publishers, particularly those publishers who only produce one or two journals. Such publishers, who may not have dedicated technical support, are less likely to be able to implement and maintain a repository supporting the full OAI-PMH. They might however be able to maintain a static repository, and participate in these wider networks in this way.

This article introduces STARGATE (Static Repository Gateway and Toolkit: Enabling small publishers to participate in OAI-PMH-based services) [4], a project funded by the Joint Information Systems Committee (JISC) and based in the Centre for Digital Library Research at the University of Strathclyde, which is undertaking an investigation of the applicability of this technology to small publishers.

[2] First published as Robertson R.J. (2006) Stargate: exploring static repositories for small publishers *Ariadne* 47 http://www.ariadne.ac.uk/issue47/robertson/ reprinted with permission.

Background

OAI-PMH grew out of an attempt by members of the e-print community to improve access to and dissemination of scholarly communication [5]. The success of the protocol is demonstrated in its implementation, not only in the software commonly used to create e-print repositories (such as Eprints, Dspace, and Fedora) but also in the growth of services that take advantage of the increased access to metadata it allows. The experimental registry at UIUC (University of Illinois at Urbana-Champaign) currently lists 987 existing repositories supporting OAI-PMH [6].

The protocol has found extensive use among data providers, not only because it facilitates the exchange of data (and so has allowed the construction of federated collections of metadata) but also because of the development of a number of specific services that use this metadata. Examples of these include: OAIster [7] - aiming to provide 'a collection of freely available, previously difficult-to-access, academically-oriented digital resources that are easily searchable by anyone' and Citebase [8] - providing 'a semi-autonomous citation index for the free, online research literature. It harvests pre- and post- prints (most author self-archived) from OAI-PMH-compliant archives, parses and links their references and indexes the metadata in a search engine'.

The Benefits and Problems of OAI-PMH Exposure

The OAI-PMH based exposure of metadata held in databases allows services and search engines to index records not otherwise visible to automated processes. For example, Google is harvesting and indexing materials from the National Library of Australia's digital collections through OAI-PMH [9]. The greater availability of metadata to search engines that this technology allows has resulted in increased visibility for scholarly works and other types of assets whose metadata had previously only been visible through a specific interface at a specific location (physical or virtual). This 'unlocking' of metadata has enabled greater access to information about articles and in many cases to copies of the articles themselves - benefiting not only the scholarly community but also the general public.

Although this process has benefited scholars and others, it has also created a problem about which version of an article is being described and linked to. The version of an article an author can provide to a repository is dependent on the copyright agreement between the author and publisher. Thus the metadata record for any given article can link to the deposited copy (pre-print or post-print), the publisher's copy, both copies, or no copy. This variety creates a difficulty for users and publishers; in that, for any given article, the metadata and link(s) to a copy of the paper which are retrieved by a search may not correspond to the formally-published peer-reviewed version (irrespective of users' rights to access the final formal version), and, even if it is the formal version, users may not necessarily have enough information to allow them to cite the article properly.

The potential problem for both publishers and academics is multiplied in that, if the final version is not linked to by the data provider (i.e. the repository), the correct citation (i.e. publishers' final version accessed through their designated provider) of a paper will not occur in higher-level services. This creates the potential for scholars to be referring to the same intellectual effort but in different instantiations - for example, there may be differences in page numbering, content, date of publication, and even author attribution.

Another difficulty is that some of the value of a journal article comes from its co-location with other articles. The focus of a journal, the progression of relevant topics in sequential issues, and the editorial selection of complementary or conflicting articles within an issue (in particular in a themed issue) is lost as any given repository (institutional or subject) will not contain an entire journal run or even a complete issue. Even within higher-level services based on many repositories, retrieving a journal issue is currently almost impossible as the basic metadata set exposed and harvested through OAI-PMH does not explicitly record the journal issue.

A Way Forward

One way to begin obviating these problems is for publishers to become involved in OAI-PMH based services by exposing their metadata. This would not only increase the visibility of the citable formal version, but would, as services provided on the basis of harvested metadata grow in sophistication, also ensure that compound records, produced by services aggregating and disambiguating metadata, include a link to the publisher's version.

Although the involvement of publishers in the interoperability framework provided by OAI-PMH was envisioned at the start of the protocol [5], take-up by the publishing community has been slow. One example of publisher participation in OAI-PMH is that of Inderscience, a publisher of journals 'in the fields of engineering and technology, management and business administration, and energy, environment and sustainable development' [10]. Inderscience worked together with a project team from EEVL (The Internet Guide to Engineering, Mathematics and Computing) to integrate metadata about their products into external services, in particular cross-referencing services. The development of the Inderscience OAI-PMH repository was in part funded by JISC as part of the Metadata and Interoperability Projects (5/03) strand.

The experience of EEVL and Inderscience, and Inderscience's ongoing participation in OAI-PMH, suggests that, in practice as well as in theory, publishers can benefit from exposing their metadata.

An Obstacle to Participation in OAI-PMH and a Proposed Solution

There are however publishers for whom establishing and maintaining such a full OAI-PMH repository may be problematic. The case study provided by EEVL on the above repository development comments that 'the generic task of configuring a web server to handle OAI-PMH requests and parsing out the arguments should involve less than a day of work for someone experienced with setting up Web servers and writing CGI scripts' [11]. Although this task may be straightforward compared to developing other Web services, for small publishers without technical support it may still remain a significant challenge.

The community developing the Open Archives Initiative has striven to make participation in OAI-PMH as easy as possible and has developed a simpler solution. This solution uses a combination of static repositories (XML files) and a static repository gateway. All the participant has to do is create a compliant XML file, place it on a Web server, and register it with a gateway. The static repository is then available for harvesting via the gateway [12].

The utility of static repositories to lower the barriers to participation has been demonstrated in the Open Language Archives Community (OLAC), which has fostered a community 'creating a worldwide virtual library of language resources by: (i) developing consensus on best current practice for the digital archiving of language resources, and (ii) developing a network of interoperating repositories and services for housing and accessing such resources' [13]. OLAC's network includes both full repositories and static repositories, and they have, alongside the OAI_DC metadata set, implemented community-specific metadata sets to extend the services they can offer. The potential value of static repositories to lower the technical barrier to participation was also highlighted as one of the key outcomes of the HaIRST Project [14].

As the name suggests, static repositories are designed for relatively static collections of metadata. The specification of the protocol, however, allows for changes to the contents of a collection, implying that the use of static repositories for more dynamic collections is certainly possible.

Static repositories may, therefore, present an apt technical solution to allow small publishers to participate in OAI-PMH based services. This use of static repositories would represent an innovative use of the technology as it is being applied to collections of metadata that change as each issue of the journal is released. The STARGATE Project is investigating the applicability of this solution.

The STARGATE Project

The project will demonstrate the applicability of OAI-PMH static repositories by creating a series of static repositories containing publisher metadata, a gateway in which publishers' static repositories are registered and exposed. It will also demonstrate the harvesting of publisher metadata, using HaIRST's ARC harvester, and cross-searching of the exposed metadata, using the EEVL Xtra service.

The project will create case studies documenting the set-up of the static repositories, the initial tools used to support the creation of these static repositories, and will critically reflect on the strengths and weaknesses of the static repositories approach to exposing publisher metadata. This reflective analysis will draw on the publishers' impressions of the processes involved and will also draw comparisons with alternative approaches to exposing publisher metadata. The outcome of this will be to make recommendations on how, and in what circumstances, publishers might choose to implement the static repositories approach.

The four journals (all from the field of Library and Information Science) participating in the project are:

- *Journal of Digital Information* (JoDI) [15], an international peer-reviewed open access journal published by Texas A&M University
- *Information Research* [16], an independently published international peer-reviewed open access journal
- *Library and Information Research* (LIR) [17], a journal with a mix of peer reviewed and practitioner articles published by the Library and Information Research Group of the Chartered Institute of Library and Information Professionals (CILIP)
- *Information Scotland* [18], a professional journal published by the Scottish Library and Information Council on behalf of CILIPS (Chartered Institute of Library and Information Professionals Scotland)

Although all of these publishers provide electronic versions of their journal, they have different publication processes and different technical support available to them. The differences between the journals (frequency of publication, method, staff involved, metadata created) allow the applicability and efficiency of static repositories to be assessed in a variety of settings. One of the publishers, Texas A&M University, is in the process of setting up its own OAI-PMH repository, which may allow a comparison between static and full repositories.

Creating static repositories for publisher metadata will not in itself resolve the difficulties with identifying consecutive articles from particular issues of a journal. It will, however, allow for searches to be restricted to a particular journal and may inform the future development of appropriate metadata elements or extensions.

Conclusion

The outcomes of this project exploring the benefits of static repositories to the publishing community will support both the greater participation of that community within the OAI community and the wider use of static repositories. Enabling small publishers of professional and peer-reviewed journals to expose their metadata increases the visibility of the citable final version and provides other repositories with a clear link to this version. Testing the flexibility of static repositories promotes their use for other, perhaps more dynamic, content such as e-books, e-learning materials and other digital resources.

The project [4] is underway and will finish at the end of May 2006.

References

1. The Open Archives Initiative Protocol for Metadata Harvesting http://www.openarchives.org/OAI/openarchivesprotocol.html
2. Chumbe, S., Macleod, R. "Developing Seamless Discovery of Scholarly and Trade Journal Resources Via OAI and RSS", Isaias, P., Karmakar, N. eds. Proceedings of the IADIS International Conference WWW/Internet 2003 Algarve, Portugal, 5-8 November 2003 Volume 2 853-856.
3. Specification for an OAI Static Repository and an OAI Static Repository Gateway http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm
4. STARGATE http://cdlr.strath.ac.uk/stargate/
5. Lagoze, C., Van de Sompel, H., "Building a low-barrier interoperability framework", JCDL '01, June 17-23, 2001, Roanoke, VA. http://www.openarchives.org/documents/jcdl2001-oai.pdf
6. Experimental OAI Registry at UIUC http://gita.grainger.uiuc.edu/registry/
7. OAIster http://oaister.umdl.umich.edu/o/oaister/
8. Citebase http://www.citebase.org/

9. National Library of Australia Digital Object Repository http://www.nla.gov.au/digicoll/oai/
10. Inderscience Publishers Ltd. http://www.inderscience.com/mapper.php?id=11
11. Kerr, L., Corlett J., Chumbe S. (2003) Case Study for the creation of an OAI repository in a small/medium sized publishers http://www.eevl.ac.uk/projects_503.htm
12. Specification for an OAI Static Repository and an OAI Static Repository Gateway http://www.openarchives.org/OAI/2.0/guidelines-static-repository.htm
13. Open Language Archives Community http://www.language-archives.org/
14. Brophy, P. HaIRST Project summative evaluation: report. Manchester: CERLIM. (2005) http://hairst.cdlr.strath.ac.uk/documents/HAIRST-Summative-Evaluation-Final.pdf
15. *Journal of Digital Information* http://jodi.tamu.edu/
16. *Information Research* http://informationr.net/ir/
17. *Library and Information Research* http://www.cilip.org.uk/specialinterestgroups/bysubject/research/publications/journal
18. *Information Scotland* http://www.slainte.org.uk/publications/serials/infoscot/contents.html

Appendix 2

Stargate Tools

Part 1: Introduction, key issues and relevant technologies

Author: R. John Robertson and Alan Dawson

Date: 28 July 2006
Version: 1.4
Produced as part of the STARGATE Project (http//:cdlr.strath.ac.uk/stargate/)

CENTRE FOR DIGITAL
LIBRARY RESEARCH
cdlr.strath.ac.uk

INTRODUCTION

These documents are a collection of tools designed to help small journal publishers make information about their articles more accessible via a common metadata harvesting protocol – the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Specifically, it considers why exposing publisher metadata to OAI-PMH-based disclosure, discovery and alerting services is beneficial, and it supports the production of OAI-PMH static repositories from an existing collection of structured information about the articles (such as meta tags in web pages).

Objectives

These tools were designed to:

1. Support and simplify the production of static repositories from existing structured metadata.
2. Demonstrate, through a selection of case studies, the use of static repositories as a method of improving access to metadata.
3. Enable the publishing community to reflect on the appropriateness of this method to expose their metadata

Intended audience

The intended audience of these tools are small journal publishers. In particular, those who only publish a few journals, do not have dedicated technical support staff, and so are unlikely to want to support a normal (i.e. dynamic) OAI-PMH repository.

Although these tools have been developed with this specific publishing community in mind, they are also of direct relevance to others who want to make collections of digital resources available via OAI-PMH but who have similar support restrictions.

Acknowledgements

These tools have been developed as part of the Stargate project (http://cdlr.strath.ac.uk/stargate/). This project was undertaken by the Centre for Digital Library Research (CDLR; http://cdlr.strath.ac.uk/) at the University of Strathclyde, in cooperation with the TechXtra service (based at Heriot Watt University; http://www.techxtra.ac.uk/) and four independent small journals in the Library and Information Science domain. The journals were:

- Journal of Digital Information (JoDI), an international peer-reviewed open-access journal published by Texas A&M University http://jodi.tamu.edu/
- Information Research, an independently published international peer-reviewed open-access journal http://informationr.net/ir/
- Library and Information Research (LIR), a journal published by the Library and Information Research Group of the Chartered Institute of Library and Information Professionals (CILIP) http://www.cilip.org.uk/specialinterestgroups/bysubject/research/publications/journal
- Information Scotland, a professional journal published by the Scottish Library and Information Council on behalf of CILIPS (Chartered Institute of Library and Information Professionals Scotland) http://www.slainte.org.uk/publications/serials/infoscot/contents.html

Stargate was funded by the Joint Information Systems Committee (JISC) and directly builds on previous work carried out by the CDLR in the HaIRST (http://hairst.cdlr.strath.ac.uk/) project and by ICBL at Heriot Watt in the PALS I project (http://www.eevl.ac.uk/projects_503.htm). The project acknowledges the input of other interested parties and services, in particular the Association of Learned and Professional Society Publishers (ALPSP).

Project Background

The technologies used in the project and their origins are more fully explained in the introductory materials which follow.

The project brought together an experienced technical team and a group of publishers representative of the type that, it was believed, might require a simpler approach to the exposure of its metadata than is offered by the full OAI repository, and yet would be sufficiently familiar with the concepts of metadata-driven information access, standards compliance and interoperability to be able to contribute to an informed evaluation of the approach adopted within the project.

The project successfully demonstrated that this approach works. Static Repositories are relatively easy to create and that the differences between fully-fledged and static OAI Repositories would have no impact on the participation of small journal publishers in OAI-based services.

KEY ISSUES AND RELEVANT TECHNOLOGIES

What is metadata? What is OAI-PMH? Why should publishers want to make their metadata available through it? Is there a less technically complex way to participate in OAI-based services? This section addresses some of these questions.

Metadata and the promotion of digital assets

The more visible a digital asset in a diverse range of information discovery services, the easier it will be to find. Even with Google's ubiquity, specialised information discovery services still exist to cater to specific user groups and support retrieval options geared to particular community needs.[3]

Such services often act as aggregators importing or dynamically accessing information about digital assets not held within the information discovery service. These services can use different types of information to retrieve assets – from the relatively unstructured, such as full-text or link analysis of the document, to the very structured, such as a MARC21 record. The more structured information used in these services is often not intrinsic to the asset but created alongside or independently of it, and referred to as metadata.

Metadata facilitates greater precision of information retrieval queries by allowing particular characteristics of the asset to be queried. Catalogues, repositories, and other databases describing digital assets store sets of metadata about each asset – in a given set of metadata the characteristics that are described can be referred to as elements or fields. For metadata to be exchanged between such systems or to be used by aggregating information-discovery services, there must be a mechanism to make this metadata and the set of metadata elements in use available, and the distinct participating services must be compatible (interoperable) to some degree. The Dublin Core Metadata Element Set (DC) is a widely-supported set of elements for describing an asset which is often a basis for more complex local element sets that allow the basic DC element set to be easily extracted to support interaction with other systems. The Open Archive Initiative – Protocol for Metadata Harvesting (OAI-PMH) is one common exchange mechanism.[4]

[3] As well as the immediate retrieval-refinement benefits that involvement in specialised services provides, visibility in such services (especially those based in the academic domain) is also likely to improve visibility via Google.

[4] Often the DC set will be supplemented or expanded to address local information management and retrieval requirements. From this metadata a limited set corresponding to 'basic' DC will then be made available to external users. As the name suggests, OAI-PMH is a harvesting based exchange system; harvesting is a process whereby external services take a copy of the metadata records.

Digital asset providers such as commercial publishers, universities, museums or libraries can use these technologies to make information about their assets usable by external information discovery services.

Dublin Core (DC)

DC, developed in 1995 by a cross-domain meeting in Dublin, Ohio, is probably the most widely used metadata standard. Its popularity is in part because the elements were selected to comprise a minimum record set for universal resource description and, as such, most other metadata formats can be mapped to DC to some degree. Its minimalist approach is designed to provide a common metadata format to support the interoperability of resource descriptions from different domains (e.g. library and museum domains). For example, records in a domain-specific element set can be exported to DC (albeit with the likely loss of some domain-specific information) and can be used alongside records from other domains to provide a higher-level search across multiple domains.

The fifteen DC Elements (fields) are:

Title	Creator	Subject
Description	Publisher	Contributor
Date	Type	Format
Identifier	Source	Language
Relation	Coverage	Rights

Table 1. The Dublin Core Metadata Element Set

(Dublin Core Metadata Initiative, 2004. Dublin Core Metadata Element Set, Version 1.1: Reference Description. http://dublincore.org/documents/dces/)

The Open Archives Initiative and OAI-PMH

The Open Archives Initiative (OAI) developed out of a meeting in 1999 in Santa Fe to develop a technical framework to support scholarly communication through the dissemination of eprints. This technical framework, OAI-PMH, has proved very successful in supporting metadata exposure and this

> is demonstrated in its implementation, not only in the software commonly used to create e-print repositories (such as Eprints, Dspace, and Fedora) but also in the growth of services that take advantage of the increased access to metadata it allows. The experimental registry at UIUC (University of Illinois at Urbana-Champaign) currently lists 987 [1215 as of August 2006] existing repositories supporting OAI-PMH (R.J. Robertson, 2006 http://www.ariadne.ac.uk/issue47/robertson/)

The technical framework for metadata harvesting that was developed by the OAI is domain neutral – supporting domains and economic models other than open-access eprints. With its origin in the ePrints community the OAI is often misconstrued as being intrinsically supportive of open-access economic models. This misunderstanding is directly addressed in the mission statement of the Open Archives Initiative, which states:

> The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication. Continued support of this work remains a cornerstone of the Open Archives program. The fundamental technological framework and standards that are developing to support this work are, however, independent of the both the type of content offered and the economic mechanisms surrounding that content, and promise to have much broader relevance in opening up access to a range of digital materials (Open Archives Initiative, n.d. http://www.openarchives.org/organization/)

OAI-PMH, as the protocol developed by the OAI to facilitate the harvesting of metadata records from digital collections, allows repositories to provide a catalogue of their content that can be accessed and harvested by external services. Examples of specialist services that can use metadata provided through OAI-PMH include OAIster, Citebase, and Connotea. It also appears that ‘generic’ search services such as Google and Yahoo can use metadata exposed in this way.

An overview of how OAI data providers and metadata harvesters work is provided in figure 1. This illustrates how the metadata records exposed by digital repositories and libraries can be harvested by services, aggregated, and used to provide resource discovery tools (search and browse facilities) to end users. The diagram also illustrates how metadata harvesters can also be data providers exposing the aggregated record set to other services.

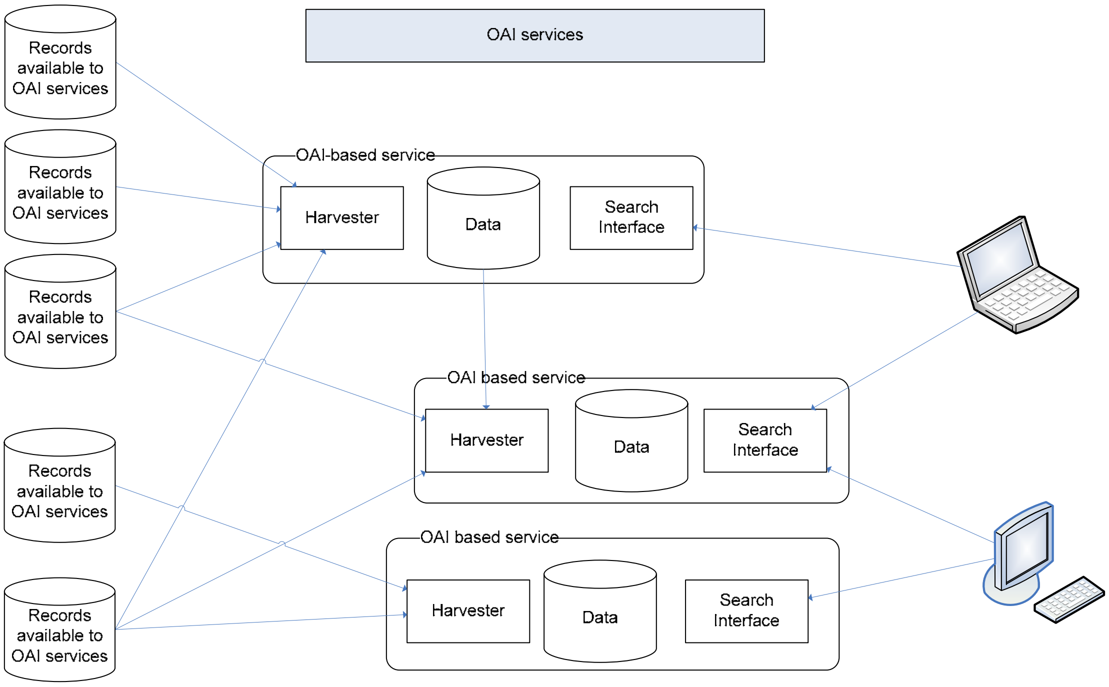

Records available to OAI services
Records available to OAI services
Records available to OAI services
Records available to OAI services
Records available to OAI services
OAI services
OAI-based service
Harvester
Data
Search Interface
OAI based service
Harvester
Data
Search Interface
OAI based service
Harvester
Data
Search Interface

Figure 1 The use of OAI-PMH by information discovery services

This mechanism enabling services to acquire records from repositories can support many metadata formats. Record sets must, however, be provided in, at least, oai_dc. Oai_dc is an XML encoding of DC, developed by the OAI as part of OAI-PMH.

Further Reading

Carl Lagoze, Herbert Van de Sompel, (2001). Building a low-barrier interoperability framework [online]. In *Joint Conference on Digital Libraries '01*, June 17-23, 2001, Roanoke, VA. http://www.openarchives.org/documents/jcdl2001-oai.pdf

Carl Lagoze et al. (eds.), (2002). The Open Archives Initiative Protocol for Metadata Harvesting. http://www.openarchives.org/OAI/openarchivesprotocol.html

Open Archives Forum (2003) OAI for Beginners - the Open Archives Forum online tutorial http://www.oaforum.org/tutorial/

Robertson R.J. (2006) Stargate: exploring static repositories for small publishers *Ariadne* 47 http://www.ariadne.ac.uk/issue47/robertson/

Evidence for the value of OAI-PMH exposure

The utility of OAI-PMH for the publishing community was initially explored in 2003 as part of the first strand of PALS Metadata and Interoperability Projects in the project, *An OAI-PMH-compliant Metadata Repository for a Specialist Publisher of E-journals*. In this project a research team from Heriot Watt University explored the development and use of an OAI-PMH repository by a medium sized publisher, Inderscience (http://www.inderscience.com) (related in the next section). Drawing on their experience in the PALS 1 project, the team at ICBL at Heriot Watt have observed the following:

> “Exposing metadata via OAI-PMH has a number of tangible benefits for publishers and content providers. Exposing metadata can:
>
> - Allow your content to be located from a large number of locations (e.g. portals, aggregators, search engines).
> - Allow aggregators to expose and promote your materials in novel ways.
> - Enhance the visibility and awareness of available resources.
> - Be a useful way to expose materials to alternative niche markets. For example, subject based services may be interested in a subset of the content from an OAI repository thus allowing exposure of their content to a more focused subject group.
> - Allow potential users to determine the relevance of resources without having to access them first.
> - Facilitate the production of interoperable services.
> - Improve the visibility of your content in search engines such as Google, Google Scholar and Yahoo.
> - Drive traffic and business to websites.”

If these are the general benefits of using OAI-PMH, there are also a number of specific problems it can help the publishing community address.[5]

> Although this process [of exposing metadata via OAI-PMH] has benefited scholars and others, it has also created a problem about which version of an article is being described and linked to. The version of an article an author can provide to a repository is dependent on the copyright agreement between the author and publisher. Thus the metadata record for any given article can link to the deposited copy (pre-print or post-print), the publisher's copy, both copies, or no copy. This variety creates a difficulty for users and publishers; in that, for any given article, the metadata and link(s) to a copy of the paper which are retrieved by a search may not correspond to the formally-published peer-reviewed version (irrespective of users' rights to access the final formal version), and, even if it is the formal version, users may not necessarily have enough information to allow them to cite the article properly.
>
> The potential problem for both publishers and academics is multiplied in that, if the final version is not linked to by the data provider (i.e. the repository), the correct citation (i.e. publishers' final version accessed through their designated provider) of a paper will not occur in higher-level services. This creates the potential for scholars to be referring to the same intellectual effort but in

[5] This section is excerpted from my article “Stargate: exploring static repositories for small publishers” *Ariadne* 47, used with permission.

> different instantiations - for example, there may be differences in page numbering, content, date of publication, and even author attribution.
>
> Another difficulty is that some of the value of a journal article comes from its co-location with other articles. The focus of a journal, the progression of relevant topics in sequential issues, and the editorial selection of complementary or conflicting articles within an issue (in particular in a themed issue) is lost as any given repository (institutional or subject) will not contain an entire journal run or even a complete issue. Even within higher-level services based on many repositories, retrieving a journal issue is currently almost impossible as the basic metadata set exposed and harvested through OAI-PMH does not explicitly record the journal issue. (R.J. Robertson, 2006)

Another example of a publisher using OAI-PMH is the Institute of Physics which has also deployed a repository to make their metadata available (http://journals.iop.org/STACKS/).

Further Reading

Robertson R.J. (2006) Stargate: exploring static repositories for small publishers *Ariadne* 47 http://www.ariadne.ac.uk/issue47/robertson/

Case Study - Implementation of an OAI-PMH Repository at Inderscience

Inderscience was one publisher that worked with Heriot Watt in their PALS 1 project. A brief case study of their experience follows.

"Inderscience is medium sized commercial journal publisher covering engineering, technology, and management & business administration. As part of a JISC/PALS funded project in 2004 the company was involved in an exercise to create an OAI–PMH compliant repository. Their motivation was commercial in nature in that they wished to "make their metadata available to aggregators and to drive more users to their full text subscription based materials". Inderscience viewed the opportunity to get as much information as possible about their content into the public domain as essential to their commercial success. Dissemination of metadata via OAI-PMH offered the possibility of users being able to discover Inderscience content from a wide variety of locations, allowing the company to play to its strengths and compete on a more level playing field with some of the larger competitors.

Experience at Inderscience revealed that it was relatively easy to set up an OAI repository, the process involving approximately 30 hours of technical time. The publisher's current Content Management System (CMS) offered a well-structured database, which required only minor modification in order to support OAI harvesting. The process involved installation of OAI data provider's software on the publisher's web site and integration with the publisher's CMS in order to populate the repository with the necessary metadata elements. The implementation required knowledge of XSLT transformations, PHP, Java servlets and MySQL. A full report on the Implementation of the Inderscience OAI Repository is available [http://www.eevl.ac.uk/casestudy.doc].

Since being involved with the project Inderscience has further developed their OAI repository, which is now publicly available [http://www.inderscience.com/mapper.php?id=15]. Their experience of implementing an OAI repository has been positive and has been instrumental in establishing partnerships with a number of content aggregators. Service providers which aggregate data from the Inderscience OAI Repository include: TechXtra [http://www.techXtra.ac.uk/], Collection of Computer Science Bibliographies [http://liinwww.ira.uka.de/bibliography] and Research Papers in Economics [http://repec.org/]."

Introduction to Static Repositories and Gateways

Although the creation of an OAI-PMH repository, as outlined in the Inderscience case study above, is 'relatively easy', it still requires a level of technical input (e.g. a "knowledge of XSLT transformations, PHP, Java servlets and MySQL") that may be beyond the everyday capabilities of smaller publishers. There is however, an easier option – a static repository.

The *Specification for an OAI Static Repository and an OAI Static Repository Gateway* was developed by the OAI to provide a simpler way of participating in OAI-PMH based services (Van de Sompel et al. (eds.), 2002 http://www.openarchives.org/OAI/2.0/guidlines-static-repository.htm). This supports, with some restrictions, a much simpler alternative technical solution to 'fully-fledged' OAI-PMH repositories.

A normal OAI repository is a database whereas a static repository is a document. A database is more flexible and powerful than a document because you can submit queries to it and extract information from it in different ways, but many people find it easier to work with documents than databases. However, if you load a document into a database then in principle it becomes possible to extract information from the document via a database query. In practice, if the document contains information in a specific predefined structure, then its content can be stored in the correct fields in a database, such as author, title, date etc.

A static repository is simply such a document with a clear structure pre-defined for this specific purpose. In more technical terms "a static repository is an XML file that is made accessible at a persistent HTTP URL. The XML file contains metadata records and repository information" (Van de Sompel et al., 2002). In other words it is a web-based structured text file accessible via a stable URL.

This file is not, in itself, able to participate in OAI-PMH-based services; it first of all needs to be registered with a static repository gateway. The gateway is a piece of software provided by a third party which is able to load documents containing metadata about journal articles (a static repository in a predefined XML format) into a structured database that can respond to OAI-PMH requests.[6] It transfers the contents of the document into the appropriate fields in a dynamic database. Thus it allows the XML file, to act as though it were a normal OAI-PMH repository. Once it has been registered in a gateway, changes in the static repository's content are detected automatically by the gateway.

Static repositories have been implemented successfully within the Open Language Archives Community (static repository gateway. http://www.language-archives.org/sr), and their use in the HaIRST project was highlighted as one of the key findings of the JISC's FAIR Programme (Peter Brophy, 2005. *HaIRST Project summative evaluation: repor*t. Manchester: CERLIM.).

The Stargate toolkit makes it possible to create a Static Repository (the XML document with the necessary predefined structure containing information about the journal articles) from a series of other documents or web pages, via an Access database.

[6] There are a number of experimental gateways available for use, and the software to create gateways is freely available. One usable gateway provided by an external party is the demonstration Gateway hosted by the OAI at the Los Alamos National Laboratories. The experimental designation primarily warns users that the service and its availability is not guaranteed.

Comparing OAI-Repositories and Static Repositories

In the instance of a small publisher producing an online journal, their initial IT infrastructure might be as follows:

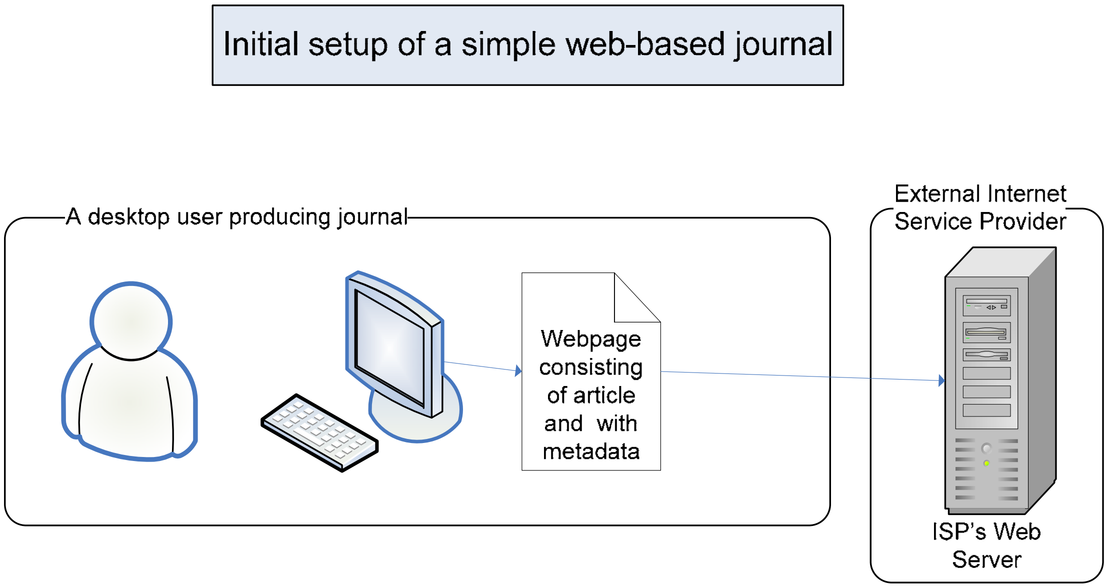

Initial setup of a simple web-based journal
A desktop user producing journal
Webpage consisting of article and with metadata
External Internet Service Provider
ISP's Web Server

Figure 2Exisitng IT infrastructure for a small journal

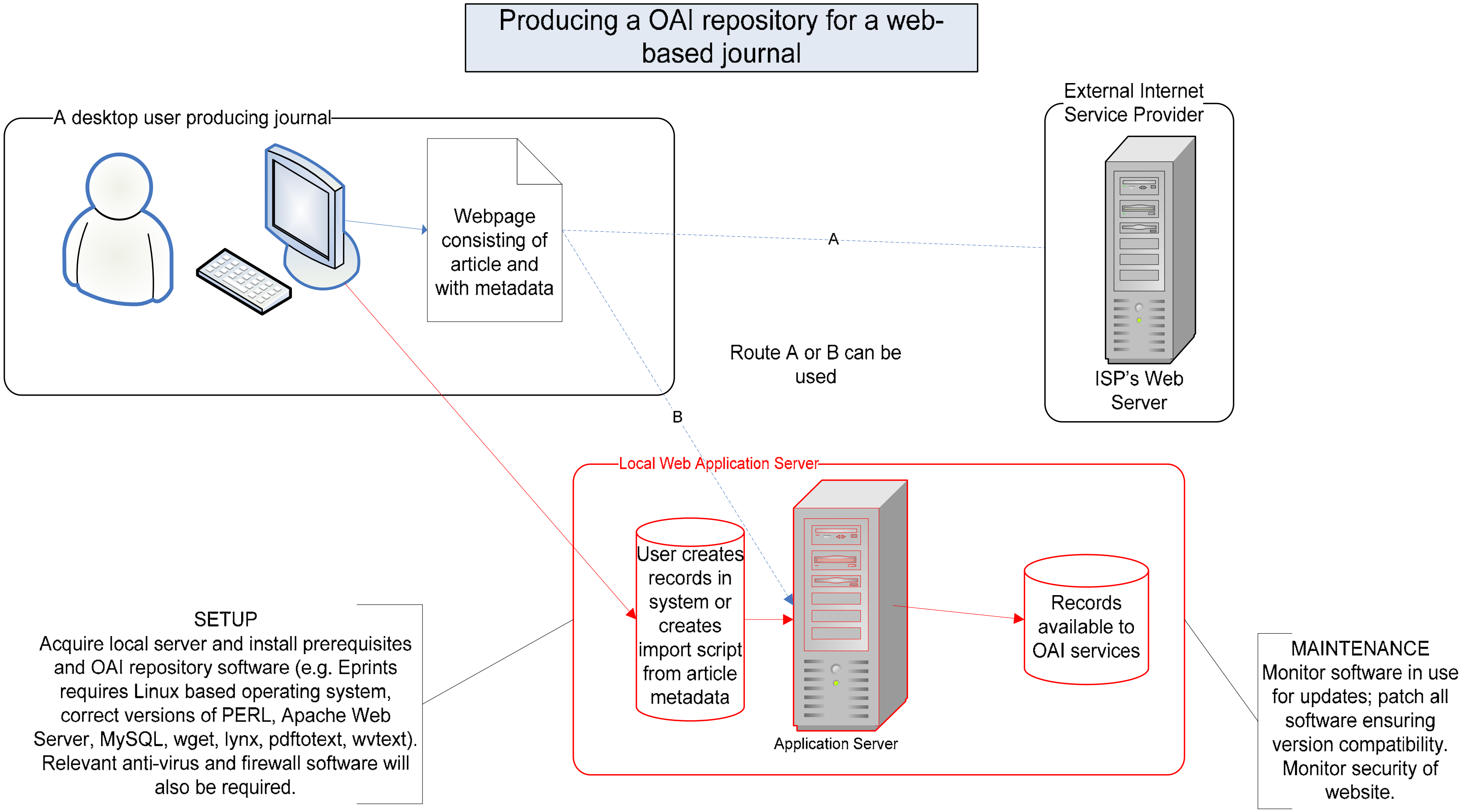
As can be seen the editor creates a webpage containing a new article for the journal (including some structured information about the article; in <meta> tags for example). They then upload it to an online folder provided by their Internet Service Provider.

For this publisher to participate in OAI-based services they need to make that structured information available in a format that OAI services can use (at least oai_dc) and in a system that an OAI harvester (the tool which collects records) can interact with. The following two diagrams illustrate the changes to the above setup that would be required to provide this availability using an OAI repository (figure 3) or using a Static Repository (figure 4).

The normal OAI repository approach is to obtain, install, configure and run a specialist piece of software that supports the OAI-PMH protocol, such as eprints.org or Dspace. 'Support the protocol' means being able to respond to OAI-PMH requests, such as 'Identify'. For example, if you type the following into a web browser: http://strathprints.strath.ac.uk/perl/oai2?verb=Identify you will get a meaningful response containing OAI 2.0 request results. In this case the OAI-PMH Identify request has been successfully interpreted by the eprints software at Strathclyde University. The technical and infrastructural requirements to do this are outlined in the diagram.

In the instance of a Static Repository a piece of software is still required, but in this case only to create XML file containing the structured information about the article. Most text editors and word-processing software can be used to do this. STARGATE chose to use an access database to create the XML file as this allowed greater flexibility in maintaining the Static Repository; it also allowed a process in which the user doesn't need to know details of the required structure of the Static Repository. The created file is considered static because it does not respond to OAI-PMH requests. It is just a file. You can not submit an OAI-PMH request to a static file.

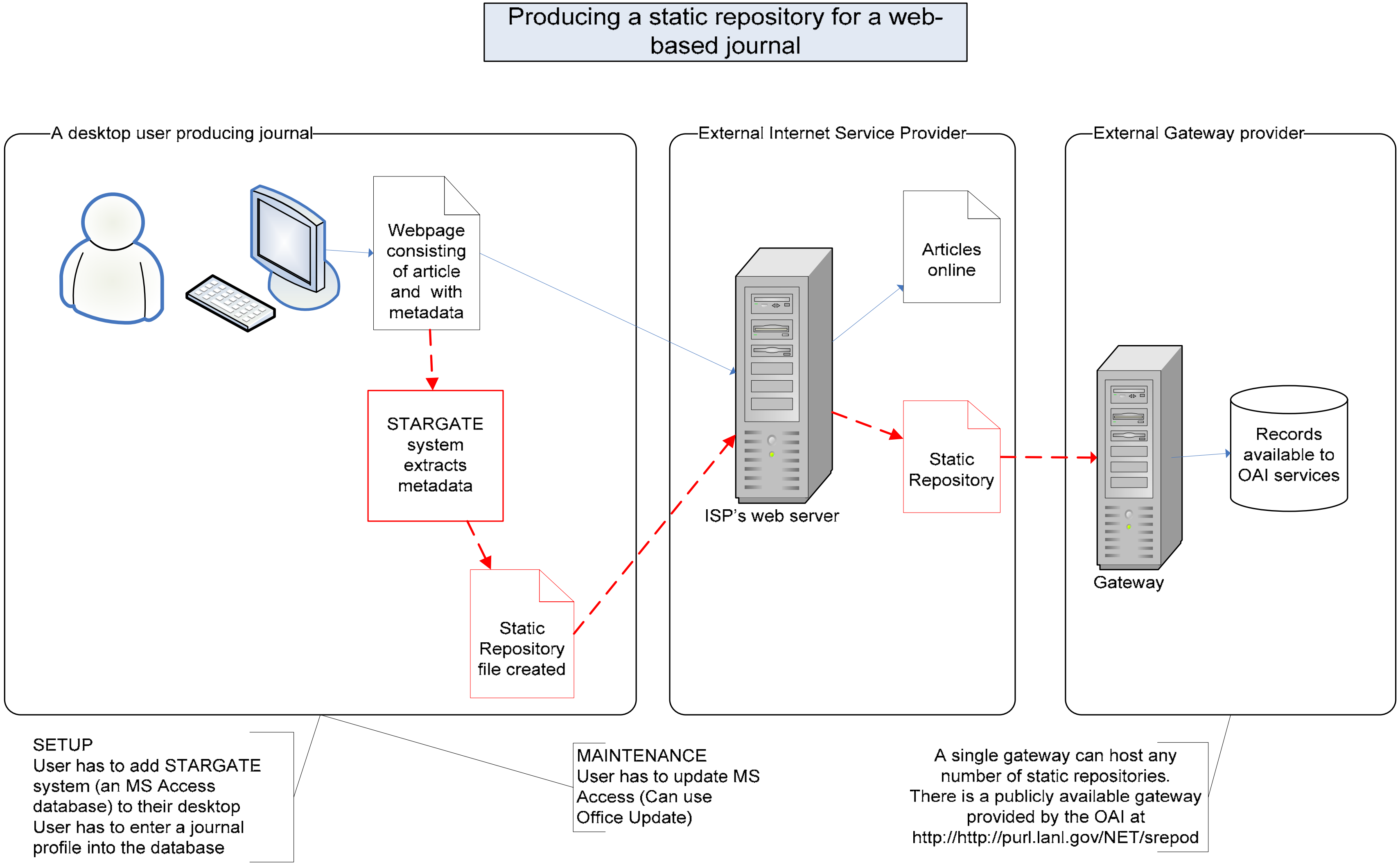
The Static Repository works because a static repository gateway can read the XML file, process the information and use it to respond to OAI-PMH requests. For example, if you type the following into a web browser: http://oaiscotland.cdlr.strath.ac.uk/services/gateway/hairst.cdlr.strath.ac.uk/repositories/jodi1.xml?verb=Identify then you will get a meaningful response about the contents of the XML file called jodi1.xml, located on a website (hairst.cdlr.strath.ac.uk), and made available through a gateway. Static Repository technology allows the infrastructure and support required to participate in OAI-based services to be shifted from the data provider (the journal) to a third party and allows a single third party gateway provider to provide intermediation for many data providers (journals).


Producing a OAI repository for a web-based journal
A desktop user producing journal
Webpage consisting of article and with metadata
External Internet Service Provider
ISP's Web Server
A
Route A or B can be used
B
Local Web Application Server
User creates records in system or creates import script from article metadata
Application Server
Records available to OAI services

SETUP
Acquire local server and install prerequisites and OAI repository software (e.g. Eprints requires Linux based operating system, correct versions of PERL, Apache Web Server, MySQL, wget, lynx, pdftotext, wvtext). Relevant anti-virus and firewall software will also be required.

MAINTENANCE
Monitor software in use for updates; patch all software ensuring version compatibility. Monitor security of website.

Figure 3 IT infrastructure to host a OAI Repository



Producing a static repository for a web-based journal

A desktop user producing journal

Webpage consisting of article and with metadata

STARGATE system extracts metadata

Static Repository file created

External Internet Service Provider

ISP's web server

Articles online

Static Repository

External Gateway provider

Gateway

Records available to OAI services

SETUP
User has to add STARGATE system (an MS Access database) to their desktop
User has to enter a journal profile into the database

MAINTENANCE
User has to update MS Access (Can use Office Update)

A single gateway can host any number of static repositories.
There is a publicly available gateway provided by the OAI at
http://http://purl.lanl.gov/NET/srepod

Figure 4 IT infrastructure to produce a Static Repository

Assumptions about users

The project's understanding of the context of small journal publishers has built on the experiences of the SAPIENS project (Scottish Academic Periodicals Implementing an Effective Networked Service, http://sapiens.strath.ac.uk/) and, through interacting with the participating journals, developed a more detailed understanding of the context of the small journal publisher community.

These tools are therefore designed with the following assumptions about the user:

1. They have at least one computer running standard business desktop software (i.e. MS Office).
2. They have access to space on a web server and can update that space themselves.
3. They already create, in some form, structured data about the articles they produce (e.g. in a database, spreadsheet, web site, or Word document). The key issue is that the information is held in some form of label and value pair (e.g. Title: Hamlet).

They can make these values available in a file that can be processed as text (e.g. plain text, or HTML format). Values stored in databases or spreadsheets could be exported into a text file or could be imported directly into the database tool using Office wizards (the tools could be used to create static repositories by manually entering information, but if structured data exists at all it is strongly advised to use data from existing mechanisms – this is less work, reduces duplication of effort, and the risk of de-synchronisation).

Appendix 3

Stargate Tools

Part 3: Issues for developing the use of OAI-based services by the publishing community

Author: R. John Robertson and Alan Dawson
Date: 28 July 2006
Version: 1.3
Produced as part of the STARGATE Project (http//:cdlr.strath.ac.uk/stargate/)

CENTRE FOR DIGITAL LIBRARY RESEARCH
cdlr.strath.ac.uk

OAI and the publishing community

Within both the publishing community and the eprints community there is an association made between open access movements and the Open Archives Initiative. This association may be due to their historic links or the extensive use of OAI-PMH by the eprints community. In its mission statement the Open Archives Initiative comments on this assumption:

> The Open Archives Initiative has its roots in an effort to enhance access to e-print archives as a means of increasing the availability of scholarly communication. Continued support of this work remains a cornerstone of the Open Archives program. The fundamental technological framework and standards that are developing to support this work are, however, independent of both the type of content offered and the economic mechanisms surrounding that content, and promise to have much broader relevance in opening up access to a range of digital materials (http://www.openarchives.org/organization/index.html)

Despite these efforts to promote the neutrality of the technology (OAI-PMH), many publishers understandably continue to be wary of the OAI, and the possible threat the predominant users of the protocol pose to their business model.

Some publishers are, however, exposing their metadata via OAI-PMH, including Inderscience and the Institute of Physics (as well as the project participants). ALPSP has also expressed interest in improving publishers' knowledge and use of the protocol. ALPSP view OAI-PMH as a key technology for publishers as it provides the opportunity for publishers to increase the visibility of their product (irrespective of the costing model the journal uses). They point out, however, that in attempting to promote the use of OAI-PMH the careful use of language is very important. Some words have very strong associations and should be used very carefully; these include: open, repository, and access.

In an emerging information environment in which the availability and quality of information about a product determines its visibility in a vast diversity of services, OAI-PMH (like other emerging standards) offers publishers a key way to be seen.

Metadata issues

For the publishing community to participate in OAI-PMH-based services there are three areas of metadata practice that should be considered. These relate to gaps in current metadata, variability in current metadata, and future developments in metadata practice. These considerations apply to publisher metadata generally and not just to metadata exposed via static repositories.

From the point of view of a service, and based on the project's experience of publishers' metadata, the metadata which is currently available is often missing some key elements. In particular, these are the URL of the article and an explicit rights statement. The URL may often be missing as the structured information is contained in the same web page as the article. The rights statement may be missing as not only is expressing rights difficult but access rights are enforced through a different mechanism (i.e. publishers do not control access to their materials through metadata). Although neither of these pieces of metadata are important when viewing the publisher's website, they become important if viewing a metadata record in an external service – they answer they question, where is this article and under what conditions can one get access to it.[7]

The second area for consideration relates to what sets of metadata are available. In the project none of the journals had chosen to implement the same set of metadata elements. All of the journals included 'title' and 'author' fields and some form of description, but after that the choice of elements varied significantly. This variability of element set choice is significant for services as they can only offer search refinement based on elements that are present across the chosen element sets of each journal. Although any metadata that is present should be visible to keyword searching, providing sophisticated search refinements requires a degree of community-wide agreement.

Following on from this the third area for consideration is that of the possible future development of a publisher-specific service. Key components of this would probably be the ability to distinguish between data types and refine searches by volume and issue. Such customised services would require the development of an alternative metadata element set (essentially an application profile containing an extended form of oai_dc). This would allow the development of community-specific specialist information discovery services. OLAC has developed and implemented this type of service and created such an application profile based on their community-specific search requirements.

Such an approach and application profile is currently under development by the eprints community. It is assumed that metadata in such an application profile would be very similar to

[7] Issues around authority control and representation of author names also fall into the category of issues to be addressed but that is an issue with metadata generally.

the sort application profile the publishing community might develop and that they would be largely congruous. This could permit the development of specialist information discovery services built on the metadata of both communities. Purely from the point of view of the metadata element set needed to enhance *access* to articles, the same application profile *could* be used by both communities. It is likely, however, that the proposed element sets will include community-specific metadata elements unrelated to enhancing access. Given that the eprints application profile is under construction already, any such development within the publishing community could benefit from building on that work.

Issues for services utilising publishers’ metadata

The need for a gateway

One precondition of the further development of this technology within the publishing community would be the creation and ongoing support of a static repository gateway for publishers’ static repositories – the repositories developed by the project are currently in a ‘research’ gateway, and their permanence cannot be assured. There are a few other existing gateways but they hold a great diversity of materials and are often also run by research groups. This is significant as there seems to be no existing Gateway that provides a ‘guaranteed’ service or presence.

Further, although the gateway itself is invisible to end users, static repositories in the same gateway are automatically listed as ‘friends’ to services. As a result, developing any form of brand or quality assurance for a group of publishers’ repositories could be difficult in a shared gateway. It is, however, unclear who could implement and maintain such a service.

Potential users

A key question for publishers thinking about exposing their metadata is who will use it and how. For open access publishers the answer is probably everybody. For commercial publishers the answers are less clear. Although OAI-PMH provides a very effective way of exposing metadata, some existing services have collection policies that clearly state that they are only interested in open access materials. Of the services involved in the project, TechXtra cross-searches all types of content, whereas OAIster and Metalis only harvest metadata about open access content. If publisher metadata were available on a large scale it is suspected that some services would begin to include it or be developed for it (both within and outwith the publishing community) – but as yet only a few such services exist.

Despite the current predilection of services towards open access content, the project participants felt strongly that there were significant information-discovery benefits for the end user in being able to access and use publishers’ metadata as well. One specific benefit envisioned is the ability to search multiple journal providers’ information at once via an interface of their choice, rather than having to search each data provider’s interface separately. Another use might be the ability to have a single interface to search for both commercial and open access materials.

Future work with publishers' metadata

In addressing publishers' metadata specifically, one longer-term avenue for a future development of the service would be to examine other ways of getting access to publishers' metadata. This project's audience were small journal publisher who, to some degree, self-published. The project was reminded that a number of small journals contract out their publication process and that another potentially effective approach to exposing their metadata would be to collaborate with such commercial publishing services. This might allow the rapid production of a seed collection of publisher metadata for services.