An Easy Option? OAI Static Repositories as a Method of Exposing Publishers' Metadata to the Wider Information Environment

R. John Robertson, Alan Dawson

Centre for Digital Library Research, Dept. of Computer & Information Sciences, University of Strathclyde
Livingstone Tower, 26 Richmond Street, Glasgow G1 1XH, UK
email: robert.robertson@cis.strath.ac.uk; alan.dawson@strath.ac.uk

Abstract

This paper introduces OAI static repository technology as a method by which publishers can expose their article-level metadata and thereby participate more fully in the wider information environment. It begins by discussing the value of exposing metadata via OAI and the potential role that static repositories might play in lowering the barriers to achieving this, and reports on the progress of a study which is evaluating the applicability and effectiveness of the static repositories approach.

Keywords: OAI-PMH; journal publisher; static repository; metadata; STARGATE; scholarly communication

1 Introduction

The scholarly communication landscape has changed significantly in the past five years. The growth of institutional and subject repositories, which aim to aid awareness of journal articles and promote the dissemination of e-prints, has made a significant impact on the visibility of such articles on the World Wide Web and has led to the development of services focusing on providing awareness of and access to the collections of articles held within them. The success of this trend is in part due to the development of a relatively simple technical protocol to support the exchange of metadata about these collections.

This protocol, the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), and its associated technologies are now well-established as mechanisms for distributing metadata about journal articles and other resource types within a networked information environment (defined as "the set of network or online services that support publishing and use of information and learning resources" [1]). Although the Open Archives Initiative (OAI) has links with the Open Access movement they are distinct entities with distinct objectives. In its mission statement the Open Archives Initiative comments on this:

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 [2].

1.1 Publisher Participation in the OAI

Within the original vision of the OAI protocol there appears to be an intent for publishers to be involved in the exchange of information within an information environment through the envisioned "Technical Umbrella for Practical Interoperability" provided by OAI-PMH [3]. The above excerpt from the current mission statement makes it explicit that it applies to any publisher, not just those operating under the Open Access approach.

The wide participation of publishers does not, however, appear to have happened to a significant degree. Although the number of OAI-based repositories of resources and metadata has increased dramatically in recent years, and the development of services based on this metadata has significantly improved the visibility and accessibility of journal articles among other resources, publishers have been slow to participate. One possible reason behind this is the historic link between OAI and Open Access, another is the widespread use of OAI-PMH to improve access to versions of articles other than the publisher's version and the perceived threat this

access poses to the publishing community's business model. A discussion of the putative reality and nature of this threat is beyond the scope of this paper; suffice to say that the OAI protocol is neutral about its use but that publishers have been reticent to expose their metadata through it.

1.2 Incentives for Publisher Participation

The launch of several high-profile services which can take advantage of OAI-PMH, such as OAIster (2001) and Google Scholar (2005), has created convenient and powerful tools for the retrieval of journal articles that are fast becoming key components in the information landscape. Furthermore, and perhaps more significantly in the longer term, repository content is increasingly being used by a multitude of diverse service providers that provide community-specific portals or aggregators, such as the Digital Library Federation's portal [4] and EEVL Xtra – a portal for the engineering community [5]. The use of OAI-PMH in these high-profile services and in the proliferation of portals and aggregators presents a compelling reason for publishers to reconsider their involvement in the emerging information environment and to expose their metadata via OAI-PMH.

An additional incentive for publishers to participate in the information environment in this manner relates to what information is currently being exposed about articles. Most of the journal articles available to users in most repositories are pre-prints or post-prints rather than the publishers' versions. The metadata exposed by repositories about these articles often corresponds to this pre- or post-print version, rather than the more citable publisher's version. Discrepancies between the metadata versions can occur for a variety of reasons, such as a change in the title or the assignment of page numbers. The most significant metadata difference for the publisher, however, is that although some institutional and subject repositories include links to the published version, many do not. As users increasingly make Google or a specialist portal their first port of call in information seeking, publishers run the risk of their versions not being visible in these services. If publishers do not expose their metadata via OAI (or its descendants) users may not be able to find articles they have published even when they are licensed to use them.

A third reason for publishers to get involved is to preserve and promote the identity and coherence of a journal issue. Within services based on existing repositories it is generally difficult, if not impossible, to browse through the current issue of a journal or to locate articles from a themed issue. The content of any given repository is likely to focus on a particular community – typically authors from a particular institution or a specific discipline rather than a particular publisher or journal. As participation in such repositories is mostly voluntary it is suggested that the probability of every author from a given issue of a journal depositing in any repository is low, and that the probability of any given service harvesting metadata about each article in a given journal issue is correspondingly lower still. As a consequence the idea of an issue of a journal is threatened – within the information environment journals are in danger of becoming solely a peer-review mechanism and a form of quality-branding. By participating in OAI-PMH-based initiatives that expose, aggregate and offer services based on article-level metadata, publishers can ensure that a complete set of metadata for each issue is available from one source. This should greatly improve the coverage and visibility of their journals in OAI-based services. Publisher engagement in the information environment through the exposure of their metadata via OAI will not only improve the profile of their journal, but also increase the value of such services.

1.3 An Example of Publisher Participation

Some publishers are aware of these issues and do expose their metadata via OAI, for example, Inderscience – "a publisher of high quality peer-reviewed international journals in the fields of engineering and technology, management and business administration, and energy, environment and sustainable development" [6]. In 2003, the Joint Information Systems Committee (JISC) funded a project to explore OAI for publishers. The project was conducted by EEVL, the resource hub for Engineering, Mathematics and Computing (based at Heriot-Watt University) in conjunction with Inderscience, and it developed an OAI repository to expose their metadata. The project demonstrated that publisher participation in OAI-PMH-based services was possible and useful – a finding confirmed by Inderscience's continuing deployment of the repository.

Given these demonstrable incentives for publishers to expose their metadata, and given that the benefits of participation have been demonstrated, it might be wondered why a further exploration of OAI-PMH for publishers is required. Further study is necessary because participation in OAI-PMH comes with a technical overhead. The development and maintenance of a repository serving metadata via OAI-PMH requires a specific technical infrastructure and an associated skill set. In EEVL's earlier study examining the use of the protocol, it was found that if the publisher's resources are well-organised, with metadata and established mechanisms for extraction, the task should not be onerous for someone experienced with setting up web servers and writing CGI

scripts [7]. Nevertheless, for publishers without immediate access to the required technical skills and infrastructure, this remains a fairly significant barrier. Small publishers producing a limited number of academic or trade journals in electronic or print formats are unlikely to be able or willing to deploy and support an OAI-compliant repository alongside their current systems.

1.4 The OAI Static Repository

Fortunately, the problem of supporting the technical and infrastructural overhead of an OAI repository is not unique to small publishers, and in response to such an overhead being an obstacle to participation, the OAI Executive and Technical Committee produced an alternative method for participation in OAI services – the static repository. The *Specification for an OAI Static Repository and an OAI Static Repository Gateway* specifies a version of the OAI protocol developed as an alternative to fully-fledged OAI-compliant repositories which offers, with some limitations, participation in the wider information environment with a greatly reduced technical overhead [8]. In essence the specification sets out an XML structure so that a data provider can put their data into a simple XML file on a web server. They then register that XML file with a static repository gateway run by a third party. The gateway is then able to intermediate the static repository – allowing the information in the static repository to be harvested by OAI service providers via the gateway. The specification places some restrictions on the content of the static repository when compared to a full OAI repository; for example, a static repository can not offer 'set' functionality, and the specification suggests a limit of around 5000 records per repository. It may be a helpful analogy to compare a full OAI-PMH repository with a dynamic web page (generated from a database) and a static repository with a static web page comprising fixed content and markup.

Static repositories have been implemented successfully within the Open Language Archives Community [9] as a means of widening participation in OAI-PMH-based systems and services, and their use in the HaIRST project (Harvesting Institutional Resources in Scotland Testbed) [10] was highlighted as one of the key pieces of learning from the JISC's FAIR Programme [11], [12]. The successful use of static repositories in these contexts, and the utility of the OAI protocol to publishers demonstrated in the EEVL Xtra project, strongly suggests that the static repository protocol could help small publishers participate more fully in the information environment. The use of static repositories to expose metadata which is added to on a regular basis (e.g. with each new issue of a journal) represents an innovative use of the technology for more dynamic applications.

2 Methodology

2.1 Overview

The remainder of this paper reports on a JISC-funded project, Stargate [13], which is investigating the applicability of the static repository specification to support the participation of small publishers in OAI-PMH-based services. The project builds on technical architecture developed in the HaIRST project recently completed at the University of Strathclyde [10]. Specifically, it is implementing a series of static repositories of publisher metadata, demonstrating the interoperability of the exposed metadata through harvesting and cross-searching via a static repository gateway, and conducting a critical evaluation of the static repository approach with publishers and service providers. At the time of writing the project is ongoing; this paper will therefore comment on the project's experience of setting up the static repositories and its findings thus far. The project is taking on the task of creating the static repositories with the intent that, after their viability has been demonstrated, the tools used can be offered to the publishers so that they can, if they wish, take over the responsibility for creating and maintaining their own static repository.

The project is working in partnership with four small publishers of e-journals in the field of information science, who have provided access to article-level metadata and will test components of a static repositories toolkit, providing feedback on the utility and effectiveness of the technology. The publishers represent a range of journal types and methods of publication. All are enthusiastic about the concept of exposing their metadata to service providers and are already exposing article-level metadata in some way, or developing the means to do so. None of the publishers currently use a database for generating their publications or storing metadata, though two of them are currently considering moving to a database-driven system. None currently has an OAI-compliant repository, and only one has the staff skills and resources necessary to develop their own repository. The project is also drawing on the experience of a number of service providers who will attempt to harvest and/or cross-search the metadata.

2.2 Method of Creating Static Repositories

There are multiple stages in the process of creating a static repository:

1. Capture or acquire metadata
2. Concatenate article-level metadata into a single file
3. Parse file and load metadata into database:
    a. Clean metadata
    b. Disaggregate metadata into elements and read into appropriate fields
4. Create OAI Static Repository from database
5. Register OAI Static Repository with gateway
6. Test the repository

Although in theory it is not essential to concatenate the metadata into a single file prior to loading it into the database, in practice this has been effective in allowing the same process to be applied to different publications.

2.3 Stage One

In this stage metadata for each journal article is captured or acquired from the publisher. This can be in the form of a file (such as a webpage) or series of files with some regular naming convention or folder structure. These files may contain just the metadata or may have the metadata embedded within other content. Metadata files have been received and parsed in HTML, RDF and plain text format.

2.4 Stage Two

In this stage the captured metadata for each article is concatenated into one file per journal so that it can be parsed into the database. The concatenation process has to ensure that there is a distinctive marker (a value or text string) which delimits the start of each article within the concatenated file. This could be pre-existent in the metadata or could be added by the process. For metadata being concatenated from multiple files the process also has to ensure that it captures any information required by the static repository that is inherent in the naming of the files or structure of the folders. It is conceivable that a journal could reach the 5000-record limit on static repositories, in which case two or more static repositories could be created for a journal.

2.5 Stage Three

In this stage the concatenated file is transformed and parsed into a database (in the Stargate project the database was developed in MS Access). As it is parsed into the database by a Visual Basic program (an Access module), the metadata records are cleaned and disaggregated. The parser reads the file looking for the marker which delimits each record. Each time it finds the marker, it creates a new record in the database and looks for the metadata element delimiters and parses them into the appropriate fields within the appropriate tables. During this disaggregation the contents of each element of information in the source file (whether in a formal metadata element set of not) is cleaned and transformed so that the version stored in the database can be output as valid XML. Some specific details of this part of the process will be commented on in the results section.

2.6 Stage Four

In this stage a second Access module reads the data stored in the database and writes it into a new file with the structure of an OAI static repository. The components of the static repository structure (XML tags and some fixed text strings) are also stored in the database to support easy editing. The OAI static repository is created on the web server at a particular location; for example, the static repository for the Information Research journal named infres1.xml is created and placed on a web-accessible folder, giving it an address of http://hairst.cdlr.strath.ac.uk/repositories/infres1.xml.

2.7 Stage Five

In this stage the newly-created static repository is registered with a static repository gateway. This process can be carried out through the gateway's user interface or it can be carried out by issuing the register command to the

gateway via http. The register command is "<Static Repository Gateway URL>?initiate=<Static Repository URL>" [8]. For the static repository example in stage four in a gateway located at http://oaiscotland.cdlr.strath.ac.uk/services/gateway intermediation would begin by issuing the command http://oaiscotland.cdlr.strath.ac.uk/services/gateway?initiate=http://hairst .cdlr.strath.ac.uk/repositories/infres1.xml .

2.8 Stage Six

After the repository has been created and registered in a gateway, it should be tested. The form of the repository can be tested using Repository Explorer (http://purl.org/net/oai_explorer) an unofficial validation tool. The content of the repository can then be tested by having it harvested – using a harvester service provides an easy method to search and review the created records as the user sees them.

2.9 Commentary on Method

The key methodological question of this approach is why a database should be used to create the static repository. If the point of a static repository is to simplify the process, using a database could be considered an unnecessary complication and seems to be a step towards the full repository approach.

In practice, a database is a useful approach, and a relatively simple Access database will suffice. Although the creation of a script or XSLT to perform the same operation could be developed, it would be more complex and would have to be extensively tailored for each publication. The database approach breaks the process into components and allows them to easily modified. Crucially, once the database has been set up it could be completely transparent to the user, as functionally it is just part of the program which transforms existing metadata into a static repository. Furthermore, the ubiquity and ease of use of Microsoft products such as Access means that making changes to the database is significantly less intimidating for non-technical users than with larger database systems such as MySQL or enterprise-level commercial products.

The chosen method for creating static repositories was recently corroborated by another project which had created static repositories in a similar manner as part of the publication process of a journal [14], [15].

3 Results

After establishing this method for the creation of static repositories, the project has successfully created them for a number of journal issues from four participating journal publishers. These journals, all drawn from the domain of information science, are:

- *Information Research* – an international peer-reviewed open access journal, independently published and edited by Professor Tom Wilson [16].
- *Journal of Digital Information* (JoDI) – an international peer-reviewed open access journal, published by Texas A&M University [17].
- *Information Scotland* – a professional journal, available in print on subscription and in electronic format on open access, published by the Scottish Library and Information Council on behalf of the Chartered Institute of Library and Information Professionals (Scotland) [18].
- *Library and Information Research* (LIR) – a mix of peer reviewed and practitioner articles, available in print on subscription and in electronic format on open access, published by the Chartered Institute of Library and Information Professionals' Library and Information Research Group [19].

These journals represent a variety of metadata creation and publishing processes. Specific issues that occurred in establishing these repositories will be highlighted in this section.

3.1 General Transformation Processes

Irrespective of the form in which the metadata is acquired, when stage three of the above method of creating static repositories is reached, the metadata is cleaned as it is parsed into the database. A subroutine of the parsing program removes specific character values in the metadata or replaces them with alternative characters. The list of values to be changed is stored in a translation table in the database for convenience of updating. The cleaning and transformation is required both to create valid XML and to provide more accessible metadata, i.e. metadata using the common versions of characters (if they are valid) or with unnecessary formatting removed.

Some examples of the transformations carried out by this subroutine are shown below. The OldText field is the existing character in the metadata, the NewText field the replacement character or the blank. Typically transformation is required to remove stray hypertext markup, to turn smart quotes into straight quotes and to deal with long dashes and ellipsis. It should be noted that many of the characters that require transformation to provide functional OAI-PMH XML are perfectly valid HTML: their presence and transformation should not be construed as a criticism of the metadata created by any of the journals.

OldText	NewText	OldText	NewText
¿	¿	“	'
		</I>	
ñ	ñ	</PRE>	
"	'	Chr(152)	
\'	'	Chr(226) & Chr(128) & Chr(34)	-
á	á	â€"	-
Chr(226) & Chr(128) & Chr(156)		’	'

Table 1: Sample metadata transformations

3.2 Information Research

"*Information Research* is a *freely available*, international, scholarly journal, dedicated to making accessible the results of research across a wide range of information-related disciplines. It is privately published" [16]. The journal is published as a series of webpages with metadata about the articles embedded within their pages. Currently this metadata is created manually; some of it by the author and some by the editor. All the metadata undergoes manual review by the editor. A sample metadata record is included below.

```
<meta http-equiv="Content-Type" content="text/html; charset=iso-8859-1" /> [...]
<title>Comunicaci&oacute;n de conocimiento. &iquest;Habilidad de los profesores
universitarios?</title>
<meta name="dc.title" content="Comunicaci&oacute;n de conocimiento.
&iquest;Habilidad de los profesores universitarios?" />
<meta name="dc.creator" content="J. Licea de Arenas, J.V. Rodr&iacute;guez, J.A.
G&oacute;mez, M. Arenas" />
<meta name="dc.subject" content="Comunicaci&oacute;n de conocimiento.
&iquest;Habilidad de los profesores universitarios?" />
<meta name="dc.description" content="Introducci&oacute;n: [...]" />
<meta name="dc.subject.keywords" content="bibliometr&iacute;a,
alfabetizaci&oacute;n en informaci&oacute;n, profesores universitarios,
Espa&ntilde;a" />
<meta name="dc.subject" content="Comunicaci&oacute;n de conocimiento.
&iquest;Habilidad de los profesores universitarios?" />
<meta name="robots" content="all" />
<meta name="dc.publisher" content="Professor T.D. Wilson" />
<meta name="dc.coverage.placename" content="global" />
<meta name="dc.type" content="text" />
<meta name="dc.identifier" scheme="ISSN" content="1368-1613" />
<meta name="dc.identifier" scheme="URI" content="http://InformationR.net/ir/11-
1/paper243.html" />
<meta name="dc.relation.IsPartOf" content="http://InformationR.net/ir/11-
1/infres111.html" />
<meta name="dc.format" content="text/html" />
<meta name="dc.language" content="es" />
<meta name="dc.rights" content="http://creativecommons.org/licenses/by-nd-nc/1.0/"
/>
<meta name="dc.date.available" content="2005-10-15" />
```

Figure 1: A sample metadata record from information research

When such records are extracted and concatenated, the html phrase "meta http-equiv" can be used as the record delimiter for parsing. This sample illustrates the presence of accents in the metadata (for example, in the authors' names) and the presence of meta tags irrelevant to the creation of static repositories (for example, <meta name="robots" content="all" />). Although the journal issue and volume number are not specifically recorded in

the mandatory metadata element set of OAI repositories (oai_dc), the project has extracted that information from the URL of the article and stores it separately in the database. This information would be a likely component of a hypothetical publisher/journal specific metadata element set. Although there is no explicit DC metadata element for storing the volume and issue number, use of the dc.relation element seems appropriate. Dc.relation is also an appropriate place to record the journal title, if it is not present in another element. The static repository record for the above example is included below.

In common with the other journals, the publisher of *Information Research* is keen to make its metadata available via OAI as this promotes greater access to its contents. In the course of the investigation it was discovered that the publishing service at Lund university already make it available through the aggregate metadata of the journals in the Directory of Open Access Journals. The project, however, feels that there is still value in creating a specific *Information Research* OAI repository.

```
<dc:title> Comunicación de conocimiento. ¿Habilidad de los profesores
universitarios? </dc:title>
<dc:creator> J. Licea de Arenas, J.V. Rodríguez, J.A. Gómez, M. Arenas
</dc:creator>
<dc:type>text</dc:type>
<dc:publisher>Professor T.D. Wilson, University of Sheffield</dc:publisher>
<dc:format>text/html</dc:format>
<dc:language>es</dc:language>
<dc:description> Introducción: [...]</dc:description>
```

Figure 2: A sample from information research's static repository

3.4 Journal of Digital Information (JoDI)

"A peer-reviewed Web journal supported by Texas A&M University Libraries" "The JoDI vision established by the editors is to be: (a) a prestige site for papers (b) a linking 'hub' for other Web services and information (c) self financing (d) low cost to the reader (e) easily accessible" [17]. The journal is published electronically using a mixture of formats (html and pdf). In JoDI's current workflow metadata is created by the author in a web form when the paper is submitted. Information from this submitted metadata is included in the article's web page as descriptive text (but not as formal metadata in the header). This information may differ from the submitted metadata if there has been a change to the paper as a result of the peer review process (for example, the paper has been given a different title). The information is not otherwise changed by the publishing team.

```
<TABLE border=2 cellpadding=5>
<TR><TH valign=top align=left>Directory</TH><TD>/Articles/incoming/Brown</TD></TR>
<TR><TH valign=top align=left>Title</TH><TD>Integrating reading and writing of
documents</TD></TR>
<TR><TH valign=top align=left>Author 1</TH><TD></TD></TR>
<TR><TH valign=top align=left>   Forename</TH><TD>Peter
John</TD></TR>
<TR><TH valign=top align=left>   Surname</TH><TD>Brown</TD></TR>
<TR><TH valign=top align=left>   Affiliation</TH><TD>University of
Exeter, UK</TD></TR>[...]
<TR><TH valign=top align=left>Author 2</TH><TD></TD></TR>
<TR><TH valign=top align=left>   Forename</TH><TD>Heather</TD></TR>
<TR><TH valign=top align=left>   Surname</TH><TD>Brown</TD></TR>
<TR><TH valign=top align=left>   Affiliation</TH><TD>University of
Exeter, UK</TD></TR> [...]
<TR><TH valign=top align=left>Theme</TH><TD>Special issue on Future Visions of
Common-use Hypertext </TD></TR>
<TR><TH valign=top align=left>Abstract</TH><TD><PRE>Computer users
[...]</PRE></TD></TR>
<TR><TH valign=top align=left>Contents</TH><TD><PRE>Abstract
1: Introduction
2: The human need
2.1: Writing-while-reading
[...] </TABLE>
```

Figure 3: A sample metadata record from journal of digital information

```
<dc:title>Integrating reading and writing of documents</dc:title>
<dc:creator>P. Brown, H. Brown</dc:creator>
<dc:publisher>Texas A&M University Libraries</dc:publisher>
<dc:date>2004</dc:date>
<dc:format>HTML</dc:format>
<dc:language>eng</dc:language>
<dc:description> Computer users [...]</dc:description>
<dc:identifier>http://jodi.tamu.edu/Articles/v05/i01/Brown</dc:identifier>
<dc:relation>Published in Journal of Digital Information Volume 5 Issue
1</dc:relation>
```

Figure 4: A sample from journal of digital information's static repository

3.5 Information Scotland

Information Scotland is the Journal of the Chartered Institute of Library and Information Professionals in Scotland (CILIPS). The electronic edition (published in html) offers open access to the journal's articles and acts as an archive of this content. It differs from the print version in that news items and announcements are not included in the electronic edition. In the production of the electronic version partial metadata is created in the html header, and each article is professionally catalogued and classified in MARC format using OCLC's Connexion tool [20] and exported into CILIPS' local resource catalogue (as a result of using Connexion a record also appears in OCLC's global catalogue: WorldCat). Rather than create the metadata from the webpage (a process demonstrated for other journals) the project used the richer metadata available via Connexion. This was exported via Connexion's qualified Dublin Core mapping. This exported file was mapped onto the appropriate database fields, imported and used to generate unqualified oai_dc as part of the creation of a static repository. Although the rich metadata exported from Connexion was inevitably reduced, the project was able to select the most appropriate elements to map onto oai_dc.

```
<link rel="SCHEMA.dc" href="http://purl.org/dc">
<meta name="DC.Title" content="Not just a 'paper exercise'">
<meta name="DC.Title.alternativeUniform" content="Information Scotland. Vol. 3,
no. 3.">
<meta name="DC.Creator.namePersonal" scheme="MEntry" content="Raven, Debby.">
<meta name="DC.Format" scheme="IMT" content="text/html">
<meta name="DC.Contributor.nameCorporate" content="Chartered Institute of Library
and Information Professionals in Scotland.">
<meta name="DC.Contributor.nameCorporate" content="School Libraries Group
(Scottish Branch)">
<meta name="DC.Contributor.namePersonal" content="Wood, Michael.">
<meta name="DC.Publisher.place" content="Hamilton :">
<meta name="DC.Publisher" content="CILIPS,">
<meta name="DC.Date.issued" scheme="MARC21-Date" content="2005">
<meta name="DC.Description" content=""Michael Wood HMI [...]"">
<meta name="DC.Description.note" content="Part of the website: SLAINTE.">
<meta name="DC.Description.note" content="Title from title screen. Description
based on contents viewed Aug. 29, 2005.">
<meta name="DC.Identifier" scheme="URI" content="http://www.slainte.org.uk/
publications/serials/infoscot/vol3(3)/vol3(3)article10.html">
<meta name="DC.Language" scheme="ISO639-2" content="eng">
<meta name="DC.Subject.classLocal" scheme="DDC" content="027.809411">
<meta name="DC.Subject.topical" scheme="LCSH" content="Libraries and schools--
Scotland.">
<meta name="DC.Relation.requires" content="Mode of access: World Wide Web.">
<meta name="DC.Relation.isPartOf" content="Information Scotland">
<meta name="DC.Type" scheme="OCLCg" content="Text data">
<meta name="DC.Type" scheme="AACR2-gmd" content="[electronic resource].">
```

Figure 5: A sample metadata record from information Scotland

```
<dc:title>Not just a 'paper exercise'</dc:title>
<dc:creator>Raven, Debby.</dc:creator>
<dc:type>Text data</dc:type>
<dc:publisher> Chartered Institute of Library and Information Professionals in
Scotland </dc:publisher>
<dc:date>2005</dc:date>
<dc:contributor> Chartered Institute of Library and Information Professionals in
Scotland. </dc:contributor>
<dc:contributor>School Libraries Group (Scottish Branch)</dc:contributor>
<dc:contributor>Wood, Michael.</dc:contributor>
<dc:format>text/html</dc:format>
<dc:language>eng</dc:language>
<dc:description> Michael Wood HMI [...]</dc:description>
<dc:identifier>http://www.slainte.org.uk/publications/serials/infoscot/vol3(3)/vol
3(3)article10.html </dc:identifier>
<dc:subject scheme="DDC">027.809411</dc:subject>
<dc:subject scheme="LCSH">Libraries and schools--Scotland.</dc:subject>
```

Figure 6: A sample from information Scotland's static repository

3.6 Library and Information Research

Library and Information Research is published by Chartered Institute of Library and Information Professionals (CILIP)'s Library and Information Research Group. It is available in print by subscription and electronically by open access. Unfortunately the journal has as yet been unable to participate further in the project's investigation so the method of metadata generation utilised in its publication is unknown. The metadata was supplied as a concatenated file in a completely regular format. This made parsing into the database very straightforward.

```
Template-type: ReDIF-Article 1.0
Author-Name: Sarah Currier
Title: Libraries and e-learning: be inspired by INSPIRAL
Abstract: In 2001, [...]
Journal: Library and Information Research News
Pages: 4-15
Volume: 26
Number: 82
Year: 2002
File-URL: http://www.lirg.org.uk/lir/pdf/article82a.pdf
Handle: ReLIS:LIR:liinre:y:2002:i:82:p:4-15
```

Figure 7: A sample from library and information research's input file

```
<dc:title>Libraries and e-learning: be inspired by INSPIRAL</dc:title>
<dc:creator>Sarah Currier</dc:creator>
 <dc:publisher> Chartered Institute of Library and Information Professionals:
Library and Information Research Group </dc:publisher>
<dc:date>2002</dc:date>
<dc:format>pdf</dc:format>
<dc:language>eng</dc:language>
 <dc:description> In 2001, [...] </dc:description>
<dc:identifier>http://www.lirg.org.uk/lir/pdf/article82a.pdf</dc:identifier>
```

Figure 8: A sample from library and information research's static repository

3.7 Metadata Consistency

In the creation of these static repositories differences between the metadata supplied by each publisher have been observed. When the metadata has been transformed and turned into static repositories these differences have largely been retained. They do, however, have implications for services using the aggregated collection of the journals' metadata. If particular search refinements were desired such a service would have to consider addressing issues such as the standardisation of two-letter and three-letter language codes, the use of the dc.format tag (for example, text/html; html; html; pdf), the use of the dc.relation tag, the presence or absence of the dc.date, dc.rights, and dc.coverage elements.

Even with only four static repositories for this project, the dc.relation tag is being used in three different ways: i) to store the URL of the relevant issue of the journal, ii) to store the title of the journal, and iii) to store the title of the journal along with the relevant volume and issue number. While it would be relatively easy, with the database methodology used by the project, to eliminate this type of inconsistency when generating the XML files for the static repositories, it is not at all clear whether modifying the metadata in this way would be acceptable to publishers, nor what is the best convention to adopt for standardisation.

If many more journals and publishers were to make use of the OAI-PMH protocol and related services, it seems inevitable that issues of metadata inconsistency will assume greater importance and may inhibit full exploitation of the benefits offered by use of the standard protocol.

4 Conclusions

Thus far Stargate has been able to demonstrate that, for a selection of publishers, multiple types of metadata – provided in a variety of methods and formats – can be transformed into OAI-PMH static repositories with relative ease in comparison with the estimated effort that would be involved in creating, populating, and maintaining a full OAI-PMH repository for the same information. It is suggested that the static repository approach does offer a flexible and low-barrier method for publishers to participate in OAI-PMH-based services which are currently shaping the information environment.

Even in the instances where a journal's metadata is included in an aggregate collection of metadata already available to OAI-PMH-based services, it is felt that there may be value in providing a specific repository for each journal. This would support the precise retrieval of a journal run and allow the entire contents of particular issues to be located easily.

The project is not yet in a position to comment on effectiveness of the OAI exposure of these journals, though this should be possible in the near future. The demonstration of cross-searching these repositories and the investigation into the effect of frequently updating static repositories are underway and these will be commented on at the conference.

Acknowledgements

The Project acknowledges the assistance of the publishers involved, and the contribution made by the EEVL Xtra (both in relating their experiences in the earlier OAI publisher project and the forthcoming cross-searching of the project's gateway).

References

[1] JISC. *Information Environment: Development Strategy 2001 – 2005 Working Draft* [online]. JISC. Available from: <http://www.jisc.ac.uk/index.cfm?name=strat_ieds0105_draft2> [cited 28 February 2006].

[2] Open Archives Initiative. *The Open Archives Initiative: Organization* [online]. Open Archives Initiative. Available from: <http://www.openarchives.org/organization/> [cited 27 February 2006].

[3] LAGOZE, C.; van de SOMPEL, H. Building a low-barrier interoperability framework [online]. **In** *Joint Conference on Digital Libraries '01*, *June 17-23, 2001, Roanoke, VA*. Available from: <http://www.openarchives.org/documents/jcdl2001-oai.pdf> [cited 27 February 2006].

[4] Digital Library Federation. *DLF OAI Portal Browse Institutions* [online]. Digital Library Federation. Available from: <http://www.hti.umich.edu/i/imls/viewcolls.html> [cited 27 February 2006].

[5] EEVL. *EEVL Xtra*, [online]. Heriot-Watt University. Available from: <http://www.eevlxtra.ac.uk/> [cited 27 February 2006].

[6] Inderscience. *Inderscience Publishers Ltd* [online]. Inderscience. Available from: <http://www.inderscience.com/> [cited 27 February 2006].

[7] CHUMBE, S.; MACLEOD, R.. Developing Seamless Discovery of Scholarly and Trade Journal Resources Via OAI and RSS. *Proceedings of the IADIS International Conference WWW/Internet 2003 Algarve, Portugal, 5-8 November 2003*. Edit. by P. Isaias, N. Karmakar. 2003, vol. 2, p. 853-856.

[8] Open Archives Initiative. *Specification for an OAI Static Repository and an OAI Static Repository Gateway* [online]. Available from: <http://www.openarchives.org/OAI/2.0/ guidelines-static-repository.htm> [cited 27 February 2006].

[9] Open Language Archives Community. *OLAC Static Repository Gateway* [online]. Available from: < http://www.language-archives.org/sr > [cited 27 February 2006].

[10] Centre for Digital Library Research. *HaIRST: Harvesting Institutional Resources in Scotland Testbed (a JISC-FAIR project)* [online]. Available from: <http://hairst.cdlr.strath.ac.uk/> [cited 28 February 2006].

[11] BROPHY, P. HaIRST Project summative evaluation: report. Manchester : CERLIM, 2005.

[12] JISC. *Focus on Access to Institutional Resources (FAIR) Programme* [online]. Available from: <http://www.jisc.ac.uk/index.cfm?name=programme_fair> [cited 28 February 2006].

[13] Centre for Digital Library Research. *STARGATE: Static Repository Gateway and Toolkit: Enabling small publishers to participate in OAI-based services* [online]. Available from: <http://cdlr.strath.ac.uk/stargate/> [cited 28 February 2006].

[14] BARCHESI, C. Archeologia e Calcolatori, nuove strategie per la diffusione di contenuti in rete sulla base dell'OAI-PMH. *Archeologia e Calcolatori,* 2005, vol. 16.

[15] BARCHESI, C. OAISistema – verso un portale OAI per gli studi sul Mediterraneo Antico [online]. **In** *Conferenza GARR 2005*. Available from: <http://www.garr.it/conf_05/slides/c_barchesi-OAI.pdf> [cited 27 February 2006].

[16] WILSON, T. D. *Information Research: an international electronic journal*. Available from: <http://informationr.net/ir/> [cited 27 February 2006].

[17] Texas A&M University Libraries. *JoDI: Journal of Digital Information*. Available from: <http://jodi.tamu.edu/> [cited 27 February 2006].

[18] Chartered Institute of Library and Information Professionals in Scotland. *Information Scotland*. Available from: <http://www.slainte.org.uk/publications/serials/infoscot/ contents.html> [cited 27 February 2006].

[19] Library and Information Research Group. *Library and Information Research*. Available from: http://www.cilip.org.uk/specialinterestgroups/bysubject/research/publications/journal> [cited 27 February 2006].

[20] OCLC. *OCLC Connexion* [online]. Available from: <http://connexion.oclc.org/ > [cited 27 February 2006].