



## Case Study for the creation of an OAI repository in a small/medium sized publishers

<i>Author(s)</i>	Linda Kerr, Jim Corlett, Santy Chumbe
<i>Last Updated</i>	17 <sup>th</sup> November 2003
<i>Version</i>	1.0
<i>Document Name</i>	Case Study for the creation of an OAI repository in a small/medium sized publishers

### Summary

This Case Study documents the creation of an OAI repository and is aimed at both conventional publishers and organisations, for example institutions and academic departments, that publish data, but who may not have considered sharing it. A brief introduction to OAI, with further references is provided.

### Contents

<b>1. Introduction</b> .....	<b>3</b>
Aims and Objectives .....	3
Acknowledgements.....	3
<b>2. OAI FAQ</b> .....	<b>4</b>
2.1 What is OAI? .....	4
2.2 What is OAI-PMH? .....	4
2.3 What is metadata?.....	4
2.4 What can you do with OAI-PMH? .....	5
2.5 Why OAI and interoperability is an issue for publishers .....	4
2.6 How do I create an OAI Repository? .....	5
2.7 Do I lose control over my data if I create an OAI repository? .....	5
2.8 How do I let people know I have an OAI repository? .....	6
<b>3. Case Study for Inderscience</b> .....	<b>8</b>
3.1 Inderscience – Company Profile .....	8
3.2 Rationale for Creation of an OAI Repository.....	8
3.3 Methodology .....	9
Report 1 : Initial publisher's database structure and management.....	10
Report 2 : Methodology and Architecture for the Inderscience's OAI Repository .....	12
3.4 OAI-PMH Harvester at EEVL.....	14
3.5 Future Developments .....	14
<b>4. References and Sources of Information</b> .....	<b>15</b>
References .....	15

Sources of Information.....	15
1. OAI and OAI-PMH.....	15
2. The JISC Information Environment .....	15

# 1. Introduction

## Aims and Objectives

The aim of this case study is to demonstrate the issues surrounding setting up a OAI repository in a small/medium sized publisher, the company's motivation for doing so, the issues involved and the outcomes, and lessons learned. The case study is an outcome of a PALS Metadata and Interoperability Project, under the *Publishers and aggregator interoperability pilots to make metadata available for distributed searching and/or harvesting* Programme.

## Acknowledgements

Further information about EEVL, JISC, PALS and the project partners can be found on their websites at:

EEVL: The Internet Guide to Engineering, Mathematics and Computing  
<http://www.eevl.ac.uk/>

Inderscience Publishers  
<http://www.inderscience.com>

JISC: The Joint Information Systems Committee  
<http://www.jisc.ac.uk/>

PALS Metadata & Interoperability Projects  
[http://www.jisc.ac.uk/index.cfm?name=programme\\_pals](http://www.jisc.ac.uk/index.cfm?name=programme_pals)

## 2. OAI FAQ

*Much of the information in this section is taken from the excellent OAI FAQs published by the Open Archives Initiative [1], and by UKOLN [2].*

### 2.1 What is OAI?

OAI stands for the **Open Archives Initiative**, which develops and promotes interoperability standards that aim to facilitate the efficient dissemination of content. The OAI endeavour is centred at Cornell University, but is widely accepted and supported by organisations such as the Digital Library Foundation (DLF) [3], the National Science Foundation (NSF) [4] and the Coalition for Networked Information (CNI) [5]. In this case study, we explore the use of OAI-PMH in creating a repository to share metadata relating to scientific journal articles. The purpose of this sharing is to broaden access to the journal articles, via third party sites.

### 2.2 What is OAI-PMH?

OAI-PMH stands for the **Open Archive Initiative Protocol for Metadata Harvesting**. It is a simple protocol that allows content providers to make available information (metadata) about their content to third parties. It supports the regular gathering of metadata from one service to another.

It is based on common underlying Web standards - HTTP, XML and XML schemas - which means that it is fairly easy to implement if you are already running a Web server.

OAI-PMH is most widely used for eprints archives, and the roots of the project are based in the eprint community. However, the concepts in the OAI interoperability framework - exposing multiple forms of metadata through a harvesting protocol - could be applied to a wide range of digital materials, for example, images or catalogue records.

### 2.3 What is metadata?

Metadata is data about data; the information that describes an object, not the object itself. A catalogue record is a metadata record. At its simplest, it is, say, the title, author and journal field. However, it can be much more complicated, depending on how much information you want to provide about the object - subject, volume number, keywords etc.

There are a number of different metadata standards or schemas. In order to provide an OAI repository, your metadata must be structured in such a way that your metadata records can be read by other systems. OAI-PMH mandates **unqualified Dublin Core** metadata. The reason for mandating the use of unqualified DC is that it provides a base level of interoperability between services, even if they know nothing about the native metadata format used by the other service.

But the OAI-PMH metadata harvesting protocol supports the notion of multiple metadata sets, allowing communities to expose metadata in formats that are specific to their applications and domains. You can exchange any metadata you like provided it is based on XML. So, for example, you can use the OAI to exchange Dublin Core (DC) metadata, IMS metadata (IEEE LOM), XrML or ODRL rights statements, etc.

For more information, see the Dublin Core Metadata Initiative web site [6].

### 2.4 Why OAI and interoperability is an issue for publishers

It is becoming increasingly important for publishers to make their data interoperable, to allow wide dissemination of content. Dissemination of metadata about content allows the resource to be located from a large number of locations. This is particularly important for smaller, specialised publishers, who are competing with large publishers such as Elsevier. Becoming interoperable with other systems goes some way to levelling the playing field. The actual content can stay on your site, but more traffic will be directed to it. More traffic leads to more usage data and better assessment of

different resources. Increasingly users are channelled to a few “main-stream” resources, with the subsequent effect this may have on the quality of research and publishing.

Terry Hulbert, of The Institute of Physics, in a presentation to the PALS conference: Delivering Content to Universities and Colleges [7] identifies that this is a way of addressing the quality issues thrown up by the “Google” culture.

The Joint Information Systems Committee (JISC) [8] set up the PALS Interoperability and Metadata Working Group to analyse the barriers to publishers’ use of metadata, and identify possible solutions. It has produced an FAQ which presents an overview of interoperability and how publishers can make their data interoperable.

## 2.5 What can you do with OAI-PMH?

OAI-PMH allows one service to ask another service for a copy of all its metadata records, or for “some” of its metadata records. “Some” is defined in terms of a **named sub-set** (known in OAI as a set), or in terms of those records modified during a particular time period.

In the terminology used by the OAI-PMH, a data provider makes data available for **gathering** and a service provider gathers that metadata and makes it available for **searching**.

In terms of the client-server model, the data provider is a server and the service provider is a client.

So, for example, a service provider could request from a data provider all metadata records in a particular subject, if that subject has been defined in the metadata. The service provider could then give its users a simple cross-search of the records from a number of data providers in a particular subject area. In practice, most service providers gather complete archives, and rely on simple searching to allow users to find the resources they require. The myOAI ([www.myoai.com](http://www.myoai.com)) service harvests a number of OAI sources and makes them available for searching. Users can locate resources only in the subject area of, say, “ocean engineering”. Much of the usefulness of an OAI archive relies on the quality of the metadata. OAI is as useful as the metadata it transports.

## 2.6 How do I create an OAI Repository?

The OAI-PMH has been designed with easy implementation in mind. Therefore, 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.

Implementing the protocol, however, involves more than simply parsing the protocol requests. Responding to protocol requests also involves accessing or extracting your metadata. If data is well-organized, already has metadata, and has established mechanisms for extracting or deriving metadata, this task should not be onerous. In the case in this case study, the work took around ten hours. Section 3.3 in this document has step-by-step guide, and links to a tutorial and sources of further information.

## 2.7 Do I lose control over my data if I create an OAI repository?

The ‘open’ in OAI doesn’t mean freely available. Data providers can choose to restrict who can gather metadata records from them based on the IP address of the service provider, or on more complex mechanisms such as HTTP Basic Authentication or SSL.

By exposing your metadata records for gathering by other services, you are allowing people to find your content without the need to visit your Web site and use your search engine. This may result in less hits on your Web site home page. However, your metadata records will typically contain the URLs of the resources held on your site. Therefore, supporting the OAI-PMH may actually result in more hits on your site - with people going direct to your resources, rather than via your home page.

Remember that you can choose to limit how much information you expose using the OAI-PMH. For example, you may choose to expose only a limited simple DC metadata record using the OAI-PMH, forcing people to visit your site if they want to see the full metadata record.

## 2.8 How do I let people know I have an OAI repository?

Once you have created an OAI repository, you can register as a data provider with the OAI. For this, you agree to make your metadata (not necessarily your content) freely available.

<http://www.openarchives.org/registration.htm#Consequences%20of%20Registration>

Once there, your repository could be picked up by one of the OAI service providers, such as myoai, or OAIster ([oaister.umdl.umich.edu/](http://oaister.umdl.umich.edu/))

Screen dump below shows a results page from OAIster, with the repositories searched on the left side of the page, and the retrieved record, with the metadata displayed.

The screenshot shows the OAIster search results page. At the top, the OAIster logo is displayed with the tagline "...find the pearls". Below the logo is a navigation bar with links for Home, Search, View Institutions, and Help. The main content area shows the search results for "industrial technology". It indicates that 24 records were found and provides options to revise the search or view the first 10 records. On the left side, there is a "Results by Institution" section listing various repositories and the number of records found in each. On the right side, the details for the first record are displayed, including the title, author/creator, year, resource type, resource format, language, notes, subject, URL, and institution.

**Results by Institution**

- DSpace at MIT: 1 record
- NASA Technical Report Server (NTRS): 6 records
- The National Science Digital Library: 8 records
- OCLC's Experimental Thesis Catalog (XTCat): 1 record
- Research Papers in Economics (RePEc): 2 records
- University of California eScholarship Repository: 1 record
- The University of Michigan, University Library, Digital Library Production Service Collections: 2 records
- Virginia Tech Electronic Thesis and Dissertation (VTETD) Collection: 2 records

**Record 1 of 24**

Title	Advanced Quality Control System for the Thermomechanical Pulping Process Improves Quality and Reduces Energy Costs for Augusta Newsprint
Author/Creator	Not Available
Year	2002-02-01
Resource Type	Office of Industrial Technologies (OIT) Forest Products BestPractices Technical Case Study (Miscellaneous)
Resource Format	BINARY
Language	English
Note	Augusta Newsprint Co. partnered with Invensys' Pacific Simulation group to implement an advanced quality control solution for the thermomechanical pulping process at their Augusta, GA plant.
Note	NREL
Subject	OFFICE OF <b>INDUSTRIAL TECHNOLOGY</b> ;BESTPRACTICES;QUALITY CONTROL;THERMOMECHANICAL PULPING PROCESS;AUGUSTA NEWSPRINT MILL;FOREST PRODUCTS;RECYCLED NEWSPRINT PLANT ; Industry
URL	<a href="http://www.osti.gov/servlets/purl/15000362-knNLwe/native/">http://www.osti.gov/servlets/purl/15000362-knNLwe/native/</a>
Institution	NASA Technical Report Server (NTRS)

Most current services are not yet set up to deal with authentication issues, and may only pick up data providers where the content is free, but some, like Scirus ([www.scirus.com](http://www.scirus.com)), provide access to both free and subscription journal articles (in this case via ScienceDirect; Scirus is owned by Elsevier).

There are also a number of portal projects in the UK that will be able to add OAI repositories as targets to their cross-searching services.

For example, the RDN Subject Portal Project is now in its implementation phase, and will develop subject portals for the UK HE and FE communities to both free and subscription content.

For more information, see the Subject Portal website [[www.portal.ac.uk/spp](http://www.portal.ac.uk/spp)] Next is a screen dump of a demonstrator page.

[last search](#) | [new search](#) | [advanced search](#) | [previous searches](#)

**Search**

**Step 1 of 2.** Enter search keywords:

**Step 2 of 2.** Select a search profile **OR** select individual collections:

<p><b>Use a Search Profile</b></p> <p><b>Default Search Profiles</b></p> <p><input type="radio"/> <b>Test2</b> ZETOC EeVl</p> <p><input checked="" type="radio"/> <b>Test1</b> COPAC SOSIG</p> <p><input type="button" value="Search"/> <input type="button" value="Reset"/></p>	<p><b>Select individual collections</b></p> <p><input type="checkbox"/> <b>ABI Inform</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>ALTIS</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>BIOME</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>BOPCRIS</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>Biz/ed</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>Bristol Biomed Image Database</b> <a href="#">description</a> </p> <p><input checked="" type="checkbox"/> <b>CAB Abstracts</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>COPAC</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>EeVl</b> <a href="#">description</a> </p> <p><input checked="" type="checkbox"/> <b>Ei Compendex</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>HUMBUL</b> <a href="#">description</a> </p> <p><input checked="" type="checkbox"/> <b>IBSS</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>IOP - Electronic Journals</b> <a href="#">description</a> </p> <p><input checked="" type="checkbox"/> <b>Inspec</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>MIMAS</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>PSIGATE</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>RAM</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>REGARD</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>SCRAN</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>SOSIG</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>SOSIG Harvester</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>XGRAIN (Demo)</b> <a href="#">description</a> </p> <p><input type="checkbox"/> <b>ZETOC</b> <a href="#">description</a> </p> <p><input type="button" value="Search"/> <input type="button" value="Reset"/></p>
--	---

There is no definitive list of service providers, although the OAI web site has a list of repositories, and the Open Archive Forum has listings of projects, services and repositories. ([http://www.oaforum.org/oaforum\\_db/index.php](http://www.oaforum.org/oaforum_db/index.php)).

## 3. Case Study for Inderscience

### 3.1 Inderscience – Company Profile

Inderscience Publishers, a company based in Geneva, Switzerland, with its Editorial Office in Olney, UK, has 25 years' experience in journal publishing. From the outset, the company's philosophy has been to map new frontiers in emerging and developing technology areas in research, industry and governance, linking with centres of excellence worldwide to provide authoritative coverage in focused and specialist fields. It aims to foster and promote innovative thinking in the sciences, management, and policy fields, seeing the need for synergy and collaboration between these fields rather than segmentation and isolation. Hence, its objectives are to build new links, networks and collaborations between these communities of thinkers, stimulating and enhancing creative and application-oriented problem-solving for society.

Its journals fall broadly into two main subject areas: engineering and technology, and management and business administration. Within these areas, there are strong subject collections – for instance, within engineering and technology, which is obviously of major interest to EEVL, there are significant titles grouped within

- the automobile collection,
- the ICT collection,
- the materials and manufacturing collection, and
- the energy, environment and sustainable development.

### 3.2 Why Inderscience wished to create an OAI repository

Commercial motivation – make their metadata available, and to drive users to their full-text materials.

Inderscience has realised a rapid expansion recently in the number of titles registered to it (well over 100, to date), and a significant number of new journals have appeared/will appear in 2003/4. All journals are available both in paper and electronically. In addition, in order to maximise access to its collections for users, and to maximise revenue for the company, Inderscience is launching an online Full Text Collection in January 2004. This will allow full searching across all published journals, with retrieval of full text documents to subscribers or pay-per-view users.

This dual approach to the marketplace at present – new journals grouped around core collections and a new online full text database – means that it is essential for the company to get as much information as possible about the new products into the public domain. Inderscience views the making available of its metadata as one means of achieving this, and of driving users to the journals and the full text material. As mentioned elsewhere, dissemination of metadata about content allows the resource to be accessed from a wide variety of locations. This should give a small publisher like Inderscience the opportunity of highlighting its strengths: users seeking information in the topic areas mentioned above should realise the depth of Inderscience's coverage of these areas by retrieving references to articles right across each particular collection.

The ability to do this freely is not to be dismissed lightly (cf RAM [Recent Advances in Manufacturing] usage on the EEVL site: a small bibliographic database, containing no full text material, but freely available, gets significant usage not only because of its subject coverage, but because it is free to access). With Inderscience, users then get the choice of becoming a subscriber to the complete full text service, or to user-defined online collections of journal titles, or they can pay-per-view on any particular article(s) required.

In this way, Inderscience, with a comparatively small marketing and publicity budget, can hope to put its products to the test on a more level playing field, as it attempts to build up its reputation for high-quality journals against more established competitors.

### **3.3 Methodology**

The methodology for creating an OAI repository at Inderscience is listed in the following two reports:

# Report 1 : Initial publisher's database structure and management

by S.Chumbe, EEVL Technical Officer, email:[santiago@macs.hw.ac.uk](mailto:santiago@macs.hw.ac.uk)

1. Introduction
2. Analysis of the Data Management
3. Analysis of the Contents
4. Analysis of the Technology
5. References and Notes

## 1. Introduction

This report ascertains the initial state of the publisher's data management system, prior to the beginning of the project, with the aim of exploring the possibilities for creating an OAI-compliant metadata repository on the publisher's site. A desirable output of this study would be the prospect of using the available database of the publisher without major redesign efforts. In this report we will try to answer questions such as: Is the current data organisation a real or suitable database for interoperability support? Is it a structured database system? Does it store all the relevant data need for OAI-PMH harvesting [1] and interoperability? What kind of database technology does the publisher have installed? Is this database enough reliable and scalable for OAI development?

This report has been produced by EEVL, the Internet guide for engineering, mathematics and computing, as part of a JISC funded metadata & interoperability project which aims to encourage the creation of publisher's metadata, which are seamless accessible and available for distributed searching and harvesting.

## 2. Analysis of Data Management

The publisher, Inderscience Publishers Ltd.[2] publishes more than 80 scientific journals, and most of their articles are relevant to EEVL. Almost 70% of the published articles are available online from the publisher's web site. The articles are stored in an SQL database and managed from a web-based Content Management System (CMS), recently implemented by the publisher. We found that the CMS is mainly oriented to support the printed production of complete journals and to allow full-text searching of their contents, without taking into account interoperability aspects nor leaving the possibility to give open access to their database to potential external aggregators and harvesters. However, because the RDBMS and the CMS were developed in-house and using open source technology, we envisage that they can be easily adapted to support OAI technology. In conclusion, Inderscience's CMS and RDBMS are able to offer a well-structured database, which only needs minor modification in order to supporting OAI harvesting.

## 3. Analysis of Content

Having identified the database systems used at Inderscience, we moved to study the contents stored in that database. Our interest is to determine if all the relevant data is already available in the database. Our criterion of selection is based in the assumption that the metadata format of the OAI repository will be based on the Dublin Core Metadata Element Set [3]. Therefore, we should make sure that the publisher's database stores the elements mentioned below.

### Dublin Core Elements used for this Project

- DC 1 = DC.Title
- DC 2 = DC.Creator (author)
- DC 3 = DC.Subject (engineering, mathematics, computing)
- DC 4 = DC.Description (abstract)
- DC 5 = DC.Publisher (Inderscience))
- DC 6 = DC.Date (Last Updated)
- DC 7 = DC.Identifier (DOI based)
- DC 8 = DC.Date Stamp (Creation date)
- DC 9 = DC.Type (single article)
- DC 10 = DC.Format (text/plain)
- DC 11 = DC.Source (Journal Code)
- DC 12 = DC.Language (English)

- DC 13 = DC.Relation (bibliographic reference: Volume, Issue No. and publication year)
- DC 14 = DC.Rights (Inderscience)

Analysis of the database content revealed that all the required metadata is already available from the publisher's database. Therefore, we may choose a methodology for implementing a software interface to that database that allows harvesting via OAI-PMH. This could be done without the need to ask the publisher to create a "static" copy of their metadata in XML format.

#### **4. Analysis of the Technology**

The CMS is the core technology developed by the IT department of the publisher. It has been developed with object-oriented libraries written in PHP and freely available as Open Source software. It also heavily relies on the MySQL RDBMS framework. The rest of the basic software is also Open Source, for example the Operating System Linux RedHat 7.3 and the web server Apache 2.3. These IT resources are enough for supporting the implementation of OAI technology and there is no need for upgrading or acquiring hardware or new software.

#### **5. References and Notes**

[1] The Open Archives Initiative Protocol for Metadata Harvesting, OAI-PMH.

URL: <http://www.openarchives.org/OAI/openarchivesprotocol.html>

[2] Inderscience Publishers Ltd.

URL: <http://www.inderscience.com>

[3] Dublin Core Metadata Elements.

URL: <http://dublincore.org/documents/dces>

## Report 2 : Methodology and Architecture for the Inderscience's OAI Repository

by S.Chumbe, EEVL Technical Officer, email:[santiago@macs.hw.ac.uk](mailto:santiago@macs.hw.ac.uk)

### Contents.

1. Introduction
2. Architecture
3. Methodology
4. References and Notes

### 1. Introduction

In this report we present the methodology and the architecture of the OAI service for Inderscience. This implementation will provide a framework for interoperability with Inderscience, by enabling metadata from their databases to be harvested and aggregated into one searchable database/interface hosted at EEVL. Taking into account that Inderscience has already a web-accessible database, which is maintained by its own CMS, we are going to propose a methodology for implementing a software interface to this web-based database that can be harvested via an OAI-PMH harvester. The repository XML format will be generated "on-the-fly" and it will be based on the Open Archives Initiative 2.0 (OAI) specification. Thus, the repository will be able to inter-operate with any OAI-compliant service.

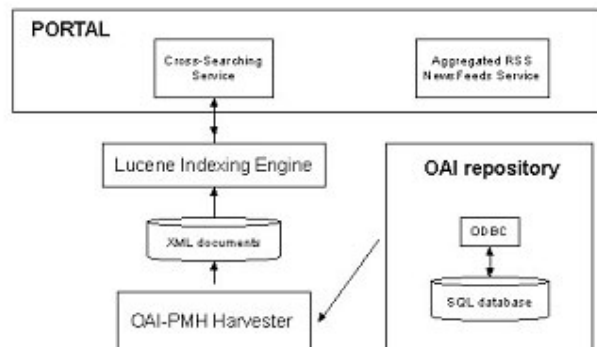
### 2. Methodology

The methodology involves the execution of the following tasks:

- Create a development area on the publisher web server and copy relevant databases for testing purposes.
- Development or adaptation and installation of a PHP/MySQL based OAI V2 data-provider software on the publisher web server.
- Integration of the data-provider software with the RDMS through the CMS, to form the OAI repository
- Help the publisher to enforce means to keep up to date the relevant databases. For instance generating guidelines to make mandatory in their database, the DC metadata elements.

### 3. Architecture

The architecture of the software framework includes an OAI-compliant repository (Data Provider) for managing the e-journal metadata, a Service Provider or harvester based on the OAI Protocol for Metadata Harvesting (OAI-PMH), and a back-end facilitator to make cross-searchable the harvested e-journals. Figure 1 shows the envisaged software architecture.



Although the software will be relatively straightforward to install, its implementation will require knowledge of XSLT transformations, Java servlets, PHP and MySQL.

#### The OAI Data Provider

The first task will be to make the existing collection of e-journals stored in the publisher site an OAI-compliant collection. Thus, we will be able to extract both the metadata and the data directly from the

publisher databases, without requiring developing additional databases. These databases will also store the necessary information for keeping track of the harvested records. We expect that the OAI-PMH Harvester can support an on-the-fly compression in order to significantly reduce the amount of data being transferred. The OAI Data-provider software will make the OAI repository accessible as a compressed XML file for the OAI harvesters. The implementation will comply with the OAI-PMH 2.0 specification, and it was inspired by the work done by U. Müller [1] at the Humboldt University of Berlin.

#### **The OAI-PMH harvester**

Similar to a web crawler, the OAI-PMH will extract metadata from the OAI repository and put it into a searchable database located on the OAI service provider, in this case EEVL. It will make use of enhanced features defined by the OAI protocol, such as the possibility to make incremental and selective harvesting.

The OAI harvester is being developed as a Java thread, which will run periodically on a web server. It will fetch the metadata from the XML file exposed by the OAI data provider, and using XSLT [2] transformations, will produce a database of XML documents.

#### **Cross-searching service**

The XML documents generated by the OAI-PMH harvester will then be fed into an indexing engine. This engine will be developed around the open source software Lucene [3]. Lucene also will provide the searching algorithms required for supporting search by query, as well as results ranking.

The user interface for searching the harvested metadata will be embedded in a bibliographic cross-searching facilitator (portlet) or channel of a portal service. This facilitator will provide the unified interface, from where any user can then search and seamlessly access the harvested metadata from different distributed e-journal archives. Currently, EEVL is actively involved in the development of the Subjects Portal Project [4], which includes the implementation of this cross-searching facilitator.

#### **4. References and Notes**

[1] Müller, U. et al, 2003. *Example of a Data Provider Implementation*. In Open Archives Forum 2003. Humboldt University of Berlin, Germany.

[2] Tidwell, D. 2001. *XSLT*. O'Reilly & Associates Publishers, San Francisco, USA

[3] Lucene Search Engine.

URL: <http://jakarta.apache.org/lucene/docs/index.html>

[4] Subject Portals Project.

URL: <http://www.portal.ac.uk/spp>

### **3.4 OAI-PMH Harvester at EEVL**

A demonstrator of the OAI-PMH of Inderscience journals is available from here:

<http://jauja.macs.hw.ac.uk/arc/index.html>

At present, linkage from the metadata record is to a journal listing page. In future developments, linkage to the journal home page or article would be desirable.

### **3.5 Future Developments**

At present, the OAI repository based at Inderscience is a demonstrator only. It is the intention that as part of the Subject Portals Project, the OAI repository will be target in the cross-search portlet. At this stage, it may be feasible to refine and improve the linking between the metadata record and the journal article.

The project has shown that it is relatively easy to set up an OAI repository. Where the challenge lies is to achieve seamless linking to content. For many publishers, interoperability is not part of their established practice, and making content available through third party sites is not yet part of their business plan. This project demonstrates that willingness is there on the part of a specialised publisher. However, the next stage for this publisher is to see the benefits of exposing their metadata in a Subject Portal context, and this, more than anything, will encourage others to take the same path.

## 4. References and Sources of Information

### References

[1] **Open Archives Initiative FAQ**

URL: <http://www.openarchives.org/documents/FAQ.html>

[2] **JISC Information Environment Architecture: OAI FAQ**

URL: <http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/faq/oai/>

[3] **Digital Library Federation**

URL: <http://www.diglib.org/>

[4] **National Science Foundation**

URL: <http://www.nsf.gov/>

[5] **Coalition for Network Information**

URL: <http://www.cni.org/>

[6] **Dublin Core Metadata Initiative**

URL: <http://dublincore.org>

[7] **PALS conference: Delivering Content to Universities and Colleges**

URL: <http://www.alpsp.org/PALS02.htm>

[8] **Joint Information System Committee**

URL: <http://www.jisc.ac.uk>

### Sources of Information

#### 1. OAI and OAI-PMH

**Open Archives Initiative FAQ**

URL: <http://www.openarchives.org/documents/FAQ.html>

**OAI for Beginners - the Open Archives Forum online tutorial**

An introduction to the Open Archives Initiative and the Protocol for Metadata Harvesting

<http://www.oaforum.org/tutorial/>

#### 2. The JISC Information Environment

**5 step guide to becoming a content provider in the JISC Information Environment**

<http://www.ariadne.ac.uk/issue33/info-environment/>

**10 minute practical guide to the Information Environment for Publishers**

<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/presentations/pals-2002-06/sld001.htm>

**JISC Information Environment Technical Architecture**

<http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/>

**Building the New Environment: A Publisher's Response**

Terry Hulbert, Institute of Physics Publishing

<http://www.alpsp.org/seminars/previous/hulPALS02.ppt>

**JISC PALS Interoperability and Metadata Working Group**

[http://www.jisc.ac.uk/index.cfm?name=wg\\_palsinter\\_home](http://www.jisc.ac.uk/index.cfm?name=wg_palsinter_home)