An easy option? OAI static repositories as a method of exposing publishers’ metadata to the wider information environment

Elpub 2006

John Robertson and Alan Dawson
Centre for Digital Library Research
University of Strathclyde
Overview

• Introduction
• Participation in OAI-based services
• The technical overhead problem and OAI-SR solution
• Stargate’s solution for publisher metadata
  – Architecture and workflow
• Metadata issues
• Results
• Conclusions and Issues
Introduction

• Stargate (Static Repository Gateway and Toolkit)
  – JISC funded short study as part of the PALS Publisher Metadata and Interoperability 2 Programme
  – Aim: to investigate the utility of OAI static repositories to enable small publishers to participate in OAI-based services within the Information Environment and beyond

• Intended audience
  – Small journals or publishers
    • minimal technical support, but some form of structured information
Acknowledgements

• Centre for Digital Library Research
  – Stargate Technical Director: Alan Dawson

• Participating journals:
  – Journal of Digital Information
  – Information Research
  – Information Scotland
  – Library and Information Research

• Harvester
  – TechXtra service at Heriot Watt
Participation in OAI-based services

• Publishers should be interested in making their metadata as available as possible – increased dissemination leads to greater impact
• Rise of services that can incorporate OAI-PMH (e.g. OAIster, Connotea, Metalis, Google Scholar/Yahoo)
• Ensuring that the final published version and the ‘correct’ metadata are present in OAI services
  – academic process is dependent on this versions
• Preserving journal identity
• Value of OAI-PMH demonstrated by use by
  – Inderscience (in PALS 1 project)
  – Institute of Physics
The technical overhead problem and OAI-SR solution

- For a lot of small publishers the required skills to create and manage an OAI repository may be too complex.
  - “knowledge of XSLT transformations, PHP, Java servelets and MySQL”

- A solution exists within the OAI Protocol in the Static Repository
  - In essence a text (xml) file on a webserver
  - The difference between SR and OAI repository is like the difference between a static versus a dynamic website
Stargate’s solution for publisher metadata

• As simple as possible actually implies still having a database
  – A database is easier to adjust as users can edit a field not wade through a XLST

• As simple as possible also implies that ubiquitous novice-friendly software should be used – MS Access and MS Word

• Some initial setup is required but if DC-like structured information exists the process is quick and significantly simpler than supporting a repository
Architecture

Structured Data → Access database → Xml file on web server

Gateway → OAI repository

Publisher

3rd Party
Locate and copy files

Concatenate files

Create Journal Profile

Create field mapping

Run 'Read Metadata files' macro

Check tables for missing or odd data

Is data in order?

YES

NO

Adjust mapping

Run 'Create OAI files' macro

Is XML valid?

YES

NO

Copy XML file to web server

Register SR with Gateway

Concatenate files or copy new article data to end of existing file

Run 'Read Metadata files' macro

Check tables for missing or odd data

Is data in order?

YES

Adjust mapping

NO

Run 'Create OAI files' macro

Is XML valid?

YES

NO

Copy XML file to web server

Creation of a new Static Repository

Updating a Static Repository
Metadata issues

• 1 unusual characters and XML
  – Foreign characters which are valid html cause XML problems. This is resolved through ‘find and replace’ during data import (managed by a table)

• 2 problems with export mappings

• 3 consistency over time / data entry errors

• 4 metadata enhancement
  – By collection level addition (e.g. journal name, publisher name)
  – By extraction (Customise script to extract volume and issue from url)
Query results in OAIster and TechXtra – provide journal specific retrieval without support for additional metadata.
Conclusions

• Findings
  – Static Repositories are an easy and effective, low tech option
  – Need for a dedicated Gateway service
  – Metadata diversity will still be an issue (elements available, no author format)

• Questions for service development
  – Journal targets: what impact would this have for users?
  – Open Access materials only or everything?
  – Possibility for dedicated service and community-specific metadata (parallel to OLAC)

• Other uses
  – OAI for a Virtual museum (http://gdl.cdlr.strath.ac.uk/springburn/)
  – Parallel methodology for EBook production (http://gdl.cdlr.strath.ac.uk/scotlandebooks.html)
Contact

Questions, comments

R. John Robertson
Centre for Digital Library Research
Department of Computer and Information Sciences
University of Strathclyde, Livingstone Tower
26 Richmond Street, Glasgow, G1 1XH, UK
Tel: +44 (0) 141 548 5854
Website: http://cdlr.strath.ac.uk/stargate
Comunicación de conocimiento. ¿Habilidad de los profesores universitarios?

Introducción: [...]
<table>
<thead>
<tr>
<th>PubId</th>
<th>PubStatus</th>
<th>PubTitle</th>
<th>Publisher</th>
<th>PubURL</th>
</tr>
</thead>
<tbody>
<tr>
<td>infres</td>
<td>active</td>
<td>Information Research</td>
<td>Professor T.D. Wilson, University of Sheffield</td>
<td><a href="http://information.r.net/">http://information.r.net/</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PubFormat</th>
<th>RecordDelim</th>
<th>FieldDelim</th>
<th>UpdateItemId</th>
</tr>
</thead>
<tbody>
<tr>
<td>html</td>
<td>newrecord</td>
<td>&lt;meta name=&quot;&quot;</td>
<td>yes</td>
</tr>
<tr>
<td>PubID</td>
<td>OldText</td>
<td>NewText</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>--------------------------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>infoscot</td>
<td>Description.abstract</td>
<td>description</td>
<td></td>
</tr>
<tr>
<td>infoscot</td>
<td>Subject.titleUniform</td>
<td>omit</td>
<td></td>
</tr>
</tbody>
</table>