Open Access Repositories Technical Considerations

Peter Millington
SHERPA Technical Development Officer





OpenDOAR

Introduction

- Δ Approaches to Setting up Repositories
 - Totally in-house Externally assisted Externally hosted
- Δ Approaches to Managing Content
 - Self-archiving Mediated service
- $\boldsymbol{\Delta}$ Things to be aware of
 - Hardware platforms & technical skills
 - Software options
 - Customisation & maintenance issues

Δ OAI-PMH

- Open Access Initiative - Protocol for Metadata Harvesting



OpenDOAR 4 6 1

Approaches to Setting up Repositories

△ Totally in-house

- Database software installed on your own servers
- Your technical staff and resources

△ Externally assisted

- Database software installed on your own servers
- Paid installation services from software providers

Δ Externally hosted

- Your repository hosted on an external server
- Consortium approach
- Institutions clubbing together
- Outsourcing to a commercial provider



In-house - Considerations

$\Delta\,$ Totally under your control

- You do all the work
- $\Delta\,$ Need a local server or space on a server
 - Sizing & capacity
 - Systems support e.g. regular backups

Δ Most repository software is open source

- Therefore cost-free
- $\Delta\,$ Availability of technical staff and resources
 - Installation and maintenance
- Δ Steep learning curve
 - Availability of external technical help



OpenDOAR 4 1

External Assistance - Considerations

Δ Also installed locally

- Usually using the same software
- Δ Division of labour
 - Installation, customisation & major upgrades
 - External service providers often the software supplier
 - Ongoing administration Local support staff

$\boldsymbol{\Delta}$ Shallower learning curve

- Fewer problems & faster installation

∆ Costs

- Software usually cost-free
- Contractual costs probably one-off



OpenDOAR 4 6 1

Consortia - Considerations

A Renefits

- Sharing hardware, effort and costs
 - Otherwise similar to an in-house repository
- Except if you are the hosting institution

$\Delta\,$ Ensuring agreement & cooperation

Δ Examples

- SHERPA-LEAP (London E-prints Access Project)
 - http://www.sherpa-leap.ac.uk/
 - Several branded interfaces to one server
- White Rose (Leeds, Sheffield, York)
 - http://eprints.whiterose.ac.uk/
 Single shared interface



OpenDOAR 4 6 1



Outsourcing - Considerations

Δ Benefits

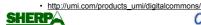
- Dedicated specialist suppliers
- No need for local technical resources quick to set up

△ Potential drawbacks

- Price?
- Constrained access e.g. 1st 24 pages of theses
- Migrating data elsewhere later

△ Example providers

- EPrints Services
 - · http://www.eprints.org/services/
- ProQuest Digital Commons



OpenDOAR

Approaches to Managing Content

$\Delta\,$ Self-archiving versus Mediated archiving - issues

- Quality of depositions metadata, uploaded full text PDFs
- Usability of the human-computer interface

$\Delta\,$ Self-archiving - Direct deposition by authors

- Variable quality of metadata e.g. Journal titles
- Requires highly usable interface
 Needs widely distributed software cost issues

Δ Mediated Archiving - Dedicated staff member

- Better standardisation of metadata & better PDFs
- Less need for an ergonomic interface, but...

 Professional software is more affordable
- Staffing costs



Technical Requirements

△ Platforms

- Operating system: Linux/Unix
- Web server: Apache
- Database: MySQL, PostgreSQL, Oracle, etc
- Scripting PERL, PHP, etc
- Java tools

$\Delta\,$ ICT staff with corresponding technical skills

Δ Software options

- PDF-making softwareRepository software

$\Delta\,$ Customisation & maintenance issues

- Keeping future upgrades in mind



OpenDOAR

PDF-Making Software

$\Delta\,$ Most eprints are deposited in PDF format

Δ Adobe Acrobat – the gold standard of PDF

- Expensive professional version not the free reader
- Appropriate for administrator not for general distribution

$\Delta\,$ Potential barrier to self-deposition

- Need to download & install PDF-making software
- Psychological adjustment You 'print' a PDF, not 'save'
- Complications Merging files for text & figures

△ Alternatives

- 2007 Microsoft Office add-in SaveAsPDF.exe
- Numerous freeware/shareware offerings
 - e.g. pdfFactory, PDFWiz, Click to Convert, etc.



OpenDOAR

Repository Software

Δ Nearly all software packages are Open Source

- Free to download plus optional paid services
- Variable quality & usability
- Variable support

Δ Most popular packages

- GNU EPrints ~30% of repositories
- DSpace ~26% of repositories
- Bepress (only hosted by ProQuest Digital Commons)
- Several dozen others

Δ Or Do-it-yourself

Fedora





GNU EPrints

$\Delta\,$ Developed by the University of Southampton

- Now http://www.eprints.org/
 Two developers + post grads: enthusiastic approach

Δ Quick to install

- LAMP platform: Linux Apache MySQL PERL
- Customisable appearance much editing of config files
- Potential upgrading problems

$\Delta \ \, \textbf{Online support}$

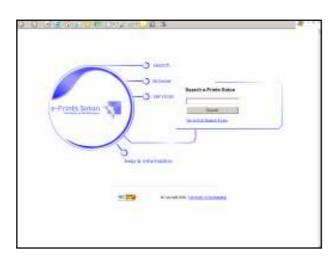
- Online documentation and Wiki
 Lively technical discussion list

$\Delta\,$ About to release Version 3

- Improved usability, especially for deposition











DSpace

$\Delta\,$ Developed by Hewlett Packard & MIT Libraries

- http://www.dspace.org/Now a true Open Source developer community

△ Installation

- Platform: Linux, Apache Ant, PostgreSQL or Oracle, Tomcat
- Most installations look the same little customisation
- Possibly few upgrade issues

$\Delta \ \, \textbf{Online Support}$

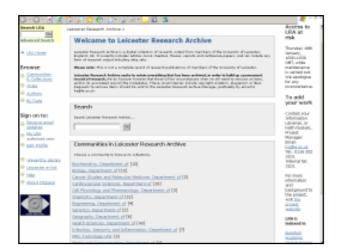
- Online documentation and Wiki
 Technical discussion list more restrained

Δ Versions

Version 1.4.1 – incremental upgrades









Fedora A Originally developed by Cornell University - http://www.fedora.info/ A One of the current buzzwords of Open Access - But few repositories claim to use Fedora A Not an out-of-the-box solution - More a Meccano® or Lego® outfit of OA components - Will it be the VHS or the Betamax of OA software? A Other Software - Often orientated towards a particular nation or language - Or totally bespoke

Customisation & Maintenance

$\boldsymbol{\Delta}$ The Need - Branding your new repository

- Giving your repository a name
- Including your institution's logo (visual identity)
- Applying your institution's full web template

∆ The Risks

- Losing track of changes done
- Problems when upgrading to new versions

$\Delta \ \ \textbf{Recommended Approach}$

- Meticulous documentation of all changes
- If possible, do a dummy run firstPrepare an 'recipe'



OpenDOAR

OAI-PMH

Δ Open Access Initiative

Protocol for Metadata Harvesting

$\boldsymbol{\Delta}$ Used to harvest information from repositories

- Search service providers
- Data miners

$\boldsymbol{\Delta}$ Maximises the expose of your research outputs

OAI-PMH is standard with GNU EPrints and DSpace

△ How it works

- Program sends HTTP request to OAI Base URL
- e.g. http://eprints.bbk.ac.uk/perl/oai2?verb=Identify
 Results returned as XML data for processing





Anu Ousstins 2	
Any Questions?	
peter.millington@nottingham.ac.uk http://www.sherpa.ac.uk/ http://www.opendoar.org/	
SHERPA E MANAGEMENT OpenDOAR	