REPOSITORIES SUPPORT PROJECT

JISC RepositoryNet

Repositories, Content Management Systems and Portals

Overview

Before examining possible institutional information architectures and the potential inclusion of repositories into that architecture, it is important to understand the functional requirements for publishing the intellectual output of the University on the web. This briefing paper discusses how repository software fits into an institution's information architecture, why repositories are quickly becoming essential information toolkits, and how they may provide the best tools for ongoing access and discovery.

Introduction

The intellectual output of the institution in the form of research papers, images, theses or similar materials is now increasingly made available on the web in order to enhance access or fulfil access requirements from funding bodies. In the modern Higher Education information environment there are now typically many systems for storing and/or managing these materials, each of which provide similar functionality. The decision of where to place this type of material is therefore often a difficult one as they could easily reside in many types of web system, from locally written systems to large scale web Content Management Systems (CMSs) or portals.

Requirements Analysis

In order to determine the most appropriate mechanism for publishing the intellectual output of the institution to the web, it is necessary to define the functional requirements typically required:

- **Open Access:** Items need to be made available online. This includes both the metadata, and the item in an appropriate digital format. Where items can not freely be made available online, the facility to restrict access is required.

- **Search engine indexing:** Sites that store materials should be accessible to crawlers and indexers in order for items to be found by conventional search engines.

- **Bibliographic metadata:** Items often come with associated metadata (title, author names, bibliographic citation etc). These details need to be stored along with the item, in a recognised metadata schema such as Dublin Core[1] or MODS[2]. The ability to 'crosswalk' between metadata schemas is also important.

- **Export functionality:** A common use of repositories is to export citation details on a per-author basis to automatically generate researcher CVs or RAE forms, or more generally to export data into bibliography manager or data analysis tools. Good export functionality can be necessary to fulfil this requirement.

- **Import functionality:** Some institutions have publication management systems or traditional websites, and want to import that data into a repository for public access.

- **Metadata harvesting:** Search system providers such as Intute Repository Search[3] and OAIster[4] use a specialist protocol known as the Open Archives Initiative Protocol for Metadata Harvesting[5] (OAI-PMH) rather than normal web crawling, which cannot accurately harvest metadata. In order for a site to be indexed by specialist search services such as these, an OAI-PMH interface is required.

- **Persistent identification:** In order to facilitate ongoing access to resources within a repository, repository software utilises tools to assign persistent identifiers. Some repository platforms make use of external identifier resolution services to ensure the identifiers would even persist if the domain name of the server or institution changed.

Integration with other systems

Repository software solutions need not be standalone systems, and can work effectively when integrated with other systems in the information architecture of an institution.

Two typical examples of repositories working along-side other systems might be:

- Using a collaborative working environment such as SharePoint™[6], authors can create, manage and apply version control to work. The work can then be published along with its metadata in the repository for final dissemination.

- If an institution has web pages in its CMS about members of staff, research CVs (publication lists) can be automatically generated from items in the repository and imported into the CMS.

Conclusion

Existing systems such as CMSs, portals, or collaborative working environments can fulfil some of the requirements identified in the list above. However they do not provide the rich tools for ongoing access and discovery found in repository software, nor do they typically provide OAI-PMH interfaces. Furthermore, the ability to export and import bibliographic metadata is often poor or non-existent. Given that these characteristics are increasingly perceived as essential to making academic output widely available over the web, a strong case can be made for integrating a repository into institutional information architectures, or even using a repository to replace existing systems, as only dedicated repository software provides this complete set of features. The key to making this decision lies in a comprehensive assessment of functional requirements and comparative analysis of existing systems with the intended repository solution. Further information on functional requirements, integration, and selecting a repository solution is available from the RSP website[7].

References & further information :

[1] Dublin Core
http://www.dublincore.org/

The Dublin Core Metadata Initiative is an open organization engaged in the development of interoperable online metadata standards that support a broad range of purposes and business models.

[2] MODS - Metadata Object Description Schema
http://www.loc.gov/standards/mods/

Metadata Object Description Schema (MODS) is a schema for a bibliographic element set that may be used for a variety of purposes.

[3] Intute Repository Search
http://www.intute.ac.uk/irs/

This service searches across 92 UK academic eprints repositories.

[4] OAIster
http://www.oaister.org/

OAIster is a union catalogue of digital resource providing access by "harvesting" descriptive metadata using OAI-PMH.

[5] OAI-PMH
http://www.openarchives.org/pmh/

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) is a low-barrier mechanism for repository interoperability.

[6] SharePoint
http://www.microsoft.com/sharepoint/

[7] Repositories Support Project
http://www.rsp.ac.uk/