

## **Breaking the Repository 'Ingest Barrier'**

**Presenter:** Scott Yeadon, Development Project Portfolio Co-ordinator, APSR

**Category:** Successful Interoperability

Many repository managers report challenges filling their institutional repositories with new scholarly resources. One of the most widely acknowledged barriers is the lack of user-friendly services that can easily connect desktop tools and applications (typically used by academics) with repositories to help automate the import, or ingest, of new resources.

In this presentation, I discuss a number of recent initiatives specifically designed to solve the repository 'ingest barrier'.

During 2007 APSR managed a series of complementary collaborative development activities undertaken by APSR partners. These activities were undertaken via the Repository Interoperability Framework (RIFF) project and addressed two major lines of development: defining specific scholarly communications workflows and developing the software services required to integrate these workflows with Fez+Fedora and DSpace. Note that workflow in the RIFF context refers to the packaging of content and metadata for deposit in a repository, transfer of the package to the repository, and unpacking the content and metadata in form manageable by the target repository.

The presentation will cover the RIFF projects and their deliverables, and demonstrate all the above services in action across several institutions. By way of description, each of the major project components are described in more detail below.

The workflow projects addressed the process of preparing and transferring scholarly materials from source applications to repositories. Four workflow projects were completed during the APSR project covering journals, conferences, images and documents. Instances of existing real-world workflows were selected and implemented: OJS journals, OCS Conferences, iSpheres image collections (an image repository and middleware developed at University of Sydney) and Scholar's Workbench documents (a scholarly authoring environment developed at ANU).

The software services component resulted in the development of a Submission Service supporting transfer from a source application to a target repository; a METS Profile to support exchange of packages between applications and repositories; and a Dissemination Service implemented as a Manakin theme editor which works with DRI-compliant repositories.

The Submission Service was developed to take a content and metadata package from some source, undertake any pre-ingest processing, and transfer the resulting package(s) to a target repository. Based around the Open Source Java-based Quartz scheduling software, the Submission Service is a repository-

independent web application that can be deployed on an existing web server or as a stand-alone service. The service is configured to run a job genre (e.g. a DSpace journal submission, a Fedora image collection submission, etc) and implementors can configure their own pre-ingest automated workflow. The service can be called from various contexts and doesn't impose a user interface or fixed workflows on the user.

To facilitate interoperability and reduce the number of packages that the Submission Service and repositories have to deal with, the Australian METS Profile was developed. The adoption of a METS Profile means standard SIP and DIPs can be used for transferring packages between source applications and repositories, and between repositories, and also be used in dissemination services. The Australian METS Profile is a three-tier model comprising a single core profile; various content profiles which inherit rules and properties from the core profile; and implementation profiles which can specify more detailed or implementation-dependent rules. For example, the Australian METS Journal Profile must conform to the core Australian METS Profile but can provide additional guidance on how to markup journal material. An implementation profile can then be developed for OJS based on the Journal Profile which outlines the requirements for packaging OJS journals, such as including a MODS descriptive metadata profile or any special requirements for handling particular versions of Galley files (e.g. HTML, PDF). Currently a METS plugin for OCS has been contributed back to the Public Knowledge Project (PKP) for possible incorporation in a future OCS release, and an OJS METS plugin will also be contributed back once complete.

The Dissemination Service resulted in the development of a Manakin theme editor based around the Moodle Chameleon theme. This allows a collection manager (in a DSpace and Fez context) to easily change a collection's look-and-feel by providing them control over the editing of the CSS and uploading of banners and other images (e.g. images on collection home pages). An additional outcome to this project was to determine the applicability of Manakin to repositories other than DSpace. To this end, work was done in partnership with the University of Queensland Fez developers to generate DRI documents within the Fez+Fedora environment. While a limited implementation it nevertheless showed that significant re-use of stylesheets and themes was achievable. An implementation for Fez+Fedora will be available in Fez 2.0.

Part of this presentation will involve a number of live demonstrations of workflow implementations across a range of source applications using Fez+Fedora and DSpace repositories as targets for archiving. The Dissemination Service will also be demonstrated showing both Fez+Fedora and DSpace examples.