

Upgration Pilot - UNSW Outcomes

Summary

Briefly summarize the goals and objectives of your pilot project.

The main goal of the Fedora 3 to 4 upgration pilot project undertaken by UNSW Library was to formulate a suitable strategy for upgrading the Library's existing Fedora 3-based repositories. A key criterion addressed by the strategy is compatibility with existing institutional data models while ensuring interoperability with related repository applications and workflows.

Development of UNSW Library's Fedora 3 to 4 upgration strategy has involved a technical assessment of Fedora 4 data model and features, followed by fedora 3 to 4 data model mapping, using an existing Library repository, [ResData](#), as a test bed.

A key output of this project is a preliminary Fedora 4 data model that is compatible with UNSW Library repositories and also aligned with the existing community Fedora 4 data model, i.e. the [PCDM](#) model. The project has also established a test Fedora 4 instance that implements the preliminary data model.

Project Details

Fedora 3 content selected, data modeling/mapping choices, tools/utilities used, final state in Fedora 4, etc.

Contents from the following two key UNSW Library Fedora 3-based repositories have been considered for this project.

- [ResData](#) - a research data management system containing over 250 records. The records describe datasets and research data management plans plus related parties (i.e. people) and activities (i.e. grants and projects). Information about people, grants and projects is sourced from other institutional databases via the data warehouse.
- [UNSWorks](#) - the institutional repository for UNSW Australia research, containing more than 12,000 publication records, including research publications such as digital theses and conference papers. The publication metadata is sourced from the Research Outputs System (ROS) with details about UNSW people and grants obtained from other UNSW enterprise systems via the data warehouse.

The main criteria for Fedora 3 to 4 data model mapping included:

- Creation of a flexible and extensible model that adequately represents current repositories such as UNSWorks and ResData while also allowing sufficient scope for future repository developments
- Re-use of existing institutional RDF ontologies, while ensuring alignment with the existing community Fedora 4 models, such as the PCDM model, and related Linked Open Data standards.
- Retention of the Fedora 3 object and datastream properties as well as the audit trail
- Implementation of customisable access constraints at both the object and collection level
- Representation of preservation events such as migration of a binary file to a preservation version
- Addressing Fedora 4's performance constraints, e.g. number of child nodes under a container

A test Fedora 4 instance for ResData has been established. A subset of the ResData records has been manually migrated to the test Fedora 4 using the aforementioned Fedora 3 to 4 migration data model.

Migration Process

Steps taken to select, analyze, and migrate data from Fedora 3 to Fedora 4, including any modifications/updates to other applications in the software stack.

Steps taken for the migration process are described below:

- Defined migration use cases
- Established a test Fedora 4 repository
- Evaluated core Fedora 4 features in comparison with related Fedora 3 features, including:
 - REST APIs
 - Versioning of records
 - Integration with external triple store
- Designed Fedora 4 data models for ResData and UNSWorks. This involved:
 - Analysis of the default Fedora 4 data model in comparison with existing ResData and UNSWorks ontologies and other related community ontologies, such as the PCDM model
 - Mapping of ResData and UNSWorks ontologies to the default Fedora 4 data model and other related community ontologies, such as the PCDM model
 - Evaluation of the Fedora 4 data model for ResData by manually migrating a subset of records to the test Fedora 4 repository
- Evaluated auxiliary Fedora 4 functions, including:
 - OAI-PMH service
 - Audit service
- Formulated a strategy for implementing the Fedora 4 REST API based on the result of evaluation of core and auxiliary Fedora 4 features.

Issues

Any issues encountered during the migration process and steps (if any) to resolve.

Issues encountered during the migration exercise include:

- Resources migrated from Fedora 3 to Fedora 4 are treated as new resources, i.e. creation date is set to the date on which the migration is complete rather than the original creation date. This is due to the data properties defined under Fedora 4 namespaces being immutable. Use of custom properties is required to enable migration of Fedora 3 default object properties, such as creation date, last modified date, and state to Fedora 4.
- Documentation about the Fedora 4 indexer configuration is inaccurate; this causes the indexer deployment to fail. This issue was resolved by troubleshooting the logs and modifying the configuration file.
- Review of the existing institutional RDF ontologies has identified some areas that require maintenance and/or could be enhanced by reusing existing standards or replacing data properties with object properties containing persistent URLs to the corresponding resources. These areas need to be explored as part of future enhancements of the institutional RDF ontologies.
- As mentioned previously, the Fedora 4 model developed has been aligned with the PCDM work. The PCDM model is very similar to the model developed for the institutional repository, UNSWorks. However the PCDM model was adapted in the following regards:
 - Inclusion of preservation migration events
 - Inclusion of separate nodes to manage access control at both the object and collection level
 - Interoperability with the ResData repository, which does not conform to a hierarchical organisation. Instead, ResData has different types of objects at an equivalent level in the repository.

Feedback

How did the migration process compare to your expectations? How could the tools, documentation, etc. be improved? Was the upgration pilot a useful exercise?

The UNSW Library Fedora 3 to 4 upgration project has provided insights into how Fedora 4 works, and will pave the way for both Fedora 3 to 4 migration and development of new Fedora 4-based repositories in future.

The documentation provided on the Fedora4 wiki was found to be generally useful and accurate, except for the Fedora 4 indexer configuration document as mentioned before. Additionally, the community Fedora 4 models, such as the PCDM model, should address performance optimisation constraints, if they are to be endorsed as Fedora 4 best practice.

Future Plans

What are your plans for continuing to migrate to Fedora 4? When do you expect to be in production?

Future work for UNSW Library Fedora 3 to 4 migration will include:

- Investigation into access control-related ontologies, such as WebACL to enable standard- based access management through the Fedora 4 data models developed by the upgration pilot.
- Implementation of a Fedora 4 REST API for integration with existing and new repository applications.
- Integrate Fedora 4 REST API with existing repository applications
- Develop tooling or customise existing Open Source tools, underpinned by the Fedora 4 models, to perform full migration of data from existing Fedora 3 repositories to Fedora 4.
- Maintain/enhance UNSW ontologies according to the Fedora 4 model developed

A new Fedora 4 repository is currently in development and testing. The migration of the first legacy repository to Fedora 4 is scheduled for Q4 2015