



Storage and Interoperability Work Package 2
Improve interoperability between Storage Resource Broker
(SRB) based environments and Fedora

Final Report

Version 1, June 2007

Lead Investigator: Andrew Treloar

Prepared by:

Andrew Treloar

andrew.treloar@its.monash.edu.au

Executive Summary

Not all large dataset environments should be, or will want to be, integrated into Fedora. The original DART bid envisaged that it would therefore also be necessary to ensure interoperability between these non-Fedora repositories and existing institutional repositories.

The objective of this work package was to improve interoperability between Storage Resource Broker (SRB) based environments and Fedora. It proposed three related activities to achieve this:

- Investigate use of Metadata Encoding and Transmission Standard (METS) to exchange archived resources between systems
- Develop authenticity management between federated systems
- Implement ability for Fedora to reference SRB-based stores as externally-managed datastreams.

In practice, none of these activities were performed in DART for the reasons described below. The DART bid actually anticipated work that was undertaken (or is about to be undertaken) by other projects.

Table of Contents

1	Introduction.....	4
2	Project Outcomes.....	5
2.1	Use of METS to exchange archived resources.....	5
2.2	Develop authenticity management between federated systems.....	5
2.3	Ability for Fedora to reference SRB-based stores.....	6
3	Archival Storage of Project Deliverables.....	7
4	Recommendations	8
5	Report Signoff.....	9

1 Introduction

The DART project worked with Fedora and SRB based repositories. It was anticipated that not all large dataset environments should be, or would want to be, integrated into Fedora. The project therefore felt it would be necessary to ensure interoperability between these non-Fedora repositories and existing institutional repositories.

This interoperability would need to be achieved in a number of different ways to accommodate different situations.

One would be the ability to exchange archived resources between systems, both over time and as researchers moved between systems. SI2 proposed to investigate using the Metadata Encoding and Transmission Standard (METS) to achieve this.

A second approach would be to develop authenticity management between federated systems, to allow for authenticated exchange of information.

The third approach was to implement the ability for Fedora to reference SRB-based stores as externally-managed datastreams. This would allow one repository to manage information stored in another. Obviously this latter approach was closely related to the work done in workpackage SI1.

2 Project Outcomes

The outcome was that none of these activities were performed in DART for the reasons described below. The DART bid actually anticipated work that was undertaken (or is about to be undertaken) by other projects. As a result, the DART Board of Management approved a scope change and the money assigned to this work package was re-assigned to central project costs.

2.1 Use of METS to exchange archived resources

Before DART could commence work on this activity, the Australian Partnership for Sustainable Repositories (APSR) announced that they were independently partnering with the National Library of Australia (NLA) on an NLA-METS profile for exchange of objects between repositories. This work is continuing over the course of 2007, and can be accessed at http://pilot.apsr.edu.au/wiki/index.php/METS_profile_development.

The ARROW and DART projects decided the most effective use of DEST funds was to contribute to this work by providing staff to serve on reference committees, attending planning workshops and submitting use-cases.

2.2 Develop authenticity management between federated systems

This work is being taken forward by two DEST funded initiatives. In both cases, it is more appropriate for the issues to be tackled at a national level.

The first initiative is the Australia Access Federation (AAF - <http://www.aaf.edu.au/>). This project will

"develop and deploy an infrastructure to facilitate trusted electronic communications and collaboration within and between higher education and research institutions both locally and internationally as well as with other organizations, in line with the NCRIS objective of providing researchers with access to an environment necessary to support world-class research. This project builds on the existing DEST funded work in the first phase of the e-Security Framework project (based at the University of Queensland) and the MAMS (Meta Access Management System project, based at Macquarie University)."

This infrastructure is a precondition to achieve this DART workpackage.

The implications of how to **use** the infrastructure in a federated data management environment will be taken up by the Australian National Data Service (ANDS) component of the Platforms for Collaboration (PfC) capability within the National Collaborative Research Infrastructure Strategy (NCRIS). For further details see http://www.ncris.dest.gov.au/capabilities/collaborative_investment_plan_platforms.htm. One of the foreshadowed ANDS components is " data re-positioning services between repositories for performance, policy and privacy reasons".

2.3 Ability for Fedora to reference SRB-based stores

While DART was underway, the Fedora developers refactored their low-level file IO modules to allow for the replacement of a conventional file store by SRB. This is now in the current release of Fedora (see <http://www.fedora.info/download/2.2/>). But this didn't address the concerns of this particular component of the SI2 workpackage.

Fedora provides a number of different types of datastream control group:

- Locally managed content:
 - Inline XML
 - Managed Content
- Externally managed content
 - External Referenced Content
 - Redirect

What DART was considering was the need for another control group: Managed External. The content of this Control Group Type would be managed (ingest/retrieve) by Fedora - but the files would be stored on a remote system. Fedora would mediate access to the remote storage.

The DART Project Architect had a number of conversations with the Fedora project architect (Sandy Payette) and the technical lead of the eSciDoc project (<http://www.escidoc.de/>) in Germany (Matthias Razum) about this problem. eSciDoc have now proposed to Fedora a solution, and have produced proof-of-concept code. This solution will meet DART's requirements. For more information, see a presentation from Matthias Razum on this solution at <http://memex.cs.cornell.edu/slider/presentation?id=fedora.info:HandlingofLargeFiles19662091>. Including this solution into a future version of Fedora is now on the development roadmap for the Fedora team.

3 Archival Storage of Project Deliverables

As the project deliverables from this workpackage were produced by other projects, DART does not have any archival responsibility. Please follow the links above for more information.

4 Recommendations

No recommendations, other than support for the activities listed above.

5 Report Signoff

It is agreed by

[Andrew Treloar](#)

That the **Final Report Document** for the [DART SI2 Improve interoperability between Storage Resource Broker \(SRB\) based environments and Fedora](#) workpackage gives a full account of the work undertaken for the DART Project.

[Andrew Treloar](#) andrew.treloar@its.monash.edu.au

- has been read and reviewed by all parties,
- shows that the work package has been completed satisfactorily,
- clearly outlines the functionality that was delivered.

Dated this [ddth](#) day of [mmmm](#) 20[yy](#)

Signed by [Andrew Treloar](#) for and on behalf of the Chief Investigator

Signed for and on behalf of DART by the Project Director [Andrew Treloar](#)