



Discovery and Access Work Package 2
Improve repository deposit rates, sharing and re-use by
improving discoverability

Final Report

Version 1, June 2007

Lead Investigator: Andrew Treloar

Prepared by:

Andrew
Treloar

andrew.treloar@its.monash.edu.au

Executive Summary

The end point of the DART e-research lifecycle model is for the various DART outputs (datasets, publications, annotations) to be discoverable.

The intention of this work package was to enable public discovery of available communication units through working integration of DART repositories, and other dataset repository with the existing National Research Discovery Service hosted by NLA at <http://search.arrow.edu.au/m> and other discipline-based discovery services as appropriate.

In fact, the work package ended up concentrating on the need to enable metadata harvesting from the annotation server that DART was enhancing, and was not able to move to testing of harvesting across multiple repositories. This will be an easy piece of future work as the necessary protocol support has now been put in place.

The National Research Discovery Service is also likely to be enhanced as part of 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.

Table of Contents

1	Introduction.....	4
2	Project Outcomes.....	5
2.1	Extend NLA RDS.....	5
2.2	Implement search interface.....	6
2.3	Integration testing.....	6
3	Archival Storage of Project Deliverables.....	7
4	Recommendations.....	8
5	Report Signoff.....	9

1 Introduction

The DART project intended to build on the discovery service already developed by the ARROW project. This service (at <http://search.arrow.edu.au/>) harvests metadata from a number of repositories (based on DSpace, Fedora, and ePrints), and then allows for searches across that metadata.

The service allows a user to search or browse across the corpus of Australia's scholarly research output which has been deposited into individual repositories at universities, the National Library of Australia and other research agencies such as Australian Policy Online.

Users can find all of the works by a researcher (even if they have moved universities), or a topic of research (from its beginnings to the current day), or the combined output of a consortium of universities. The Discovery Service currently copies records that describe research stored in university repositories, digital libraries, research centres and institutes. The records are aggregated to facilitate further discovery.

The intention of this work package was to move beyond research publications into datasets and annotations.

2 Project Outcomes

This work package achieved a number of its goals, but not all of them.

2.1 Extend NLA RDS

Conversations with the NLA indicated that their strong preference was not to extend the NLA harvester to work with new and different repositories (as foreshadowed in the DART bid). Instead (and quite reasonably) they wanted to use open standards for the communication between harvester and repository. The main standard in this area is the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH - <http://www.openarchives.org/OAI/openarchivesprotocol.html>).

This meant that DART needed to create or find OAI interfaces to the following systems:

- SRB
 - After some investigation, it was discovered that another project had already implemented an OAI interface to the SRB. See <http://www.lib.purdue.edu/research/oaisrb/> for details and the source.
- Fedora
 - OAI functionality was introduced to Fedora early on, but during the course of the DART project, the Fedora team refactored the OAI functionality as a separate OAI provider. Introduced as part of the new Fedora Service Framework, the Fedora OAI Provider Service is a stand-alone web service application that is a highly configurable OAI provider. The PROAI service can be set up to harvest any type of datastream or dissemination of digital objects in a Fedora repository. It also supports OAI sets. This new OAI service provides much more functionality and better performance. The NLA RDS was already harvesting from ARROW repositories using this software.
- Annotation server
 - The DART annotation server that was being extended in work package AA2 did not support OAI-PMH. Accordingly, DART decided to transfer some of the funds from DA2 to AA2 to enable this to be added. The OAI software was based on the open source OAICAT toolset from OCLC. For more details on the DART work, please see the relevant section of the DART website at <http://dart.edu.au/workpackages/aa/aa2.html>.

2.2 Implement search interface

Early on in the DART project, the NLA expanded the range of content types in their Research Discovery Service to include initial support for datasets (thus meeting the second objective in this work package).

Late in 2006, it became clear that the Australian National Data Service (ANDS) component of the Platforms for Collaboration (PfC) capability within the National Collaborative Research Infrastructure Strategy (NCRIS) would be looking at possible enhancements to the NLA RDS. For further details see

http://www.ncris.dest.gov.au/capabilities/collaborative_investment_plan_platforms.htm.

One of the foreshadowed ANDS components is "national discovery services across the network of digital repositories and the collection registry". It is highly likely that this will pick up some of the types of data sources that this DART work package envisaged.

2.3 Integration testing

Unfortunately, the OAI enabling of the annotation server was completed too late for us to test its integration with the NLA RDS. However, given that the OAI data providers (the dataset, repository and annotation servers) and the OAI service provider (the NLA RDS) are using the same protocol, few problems are anticipated.

3 Archival Storage of Project Deliverables

Refer to the AA2 workpackage on the DART website (<http://dart.edu.au/workpackages/aa/aa2.html#software>) for the OAI-PMH extensions to the annotation software.

4 Recommendations

That the new Australian National Data Service implement the standard OAI-PMH protocols for its discovery service. This will enable it to build on the work already done by the DART and other projects.

5 Report Signoff

It is agreed by

[Andrew Treloar](#)

That the **Final Report Document** for the [DART DA2: Improve repository deposit rates, sharing and re-use by improving discoverability](#) 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](#)