

Ready or not here it comes: Australian institutional research repository data readiness surveys 2010

Caroline Drury,
druryc@usq.edu.au,
University of Southern
Queensland

Peter Sefton,
pt@ptsefton.com,
University of Southern
Queensland

Kate Watson,
kwatson@usc.edu.au,
University of the Sunshine
Coast

Open Repositories, 2010, Madrid, Spain

Abstract

Australian institutional research repositories are now facing a new challenge: datasets and associated metadata. With prior focus predominantly on research outputs, repository managers are now involved in a new phase of repository re-purposing – curation of datasets and associated metadata, and provision of this metadata to a national data commons through ANDS (Australian National Data Service). Through a series of surveys conducted by the national repository support service, CAIRSS (the CAUL Australian Institutional Repository Support Service), this paper examines the research data challenges facing research repository managers, levels of institutional research data identification, and the readiness of traditional institutional research repositories to either curate or work alongside this data.

Keywords

institutional repository; data; data curation; data archives; metadata; community; ANDS; CAIRSS; Australian Data Commons.

Introduction

This paper examines the current state of Australian Institutional Repositories as the community gears up for a new imperative: gathering research data and placing the descriptions of those into repositories and registries, which then contribute to a national data commons. Government funding in the form of seed money to establish repositories and in repository-dependent funds were both key drivers¹ and the same is true for data. ANDS² (Australian National Data Service) is a large federal government funded project having a significant impact on the repository landscape. ANDS aims to influence national policy in the area of data management in the Australian research community, along with informing best practice for the curation of data, and transforming the disparate collections of research data around Australia into a cohesive collection of research resources. Australian institutional repositories have long been concerned with best practice in document curation primarily of research outputs,

¹ M. A Kennan and D. A Kingsley, 'The state of the nation: A snapshot of Australian institutional repositories', *First Monday*, 14 (2009).

² Australian National Data Service, 'Australian National Data Service (ANDS)', 2010 <<http://ands.org.au/>> [accessed 19 May 2010].

however with the advent of ANDS, they face a new challenge for their repositories. This presentation will summarize the current Australian institutional repository state-of-play, and readiness to curate and archive datasets.

CAIRSS³ (CAUL Australian Institutional Repository Support Service) has been running since March 2009, and has drawn together a network of repository managers built throughout earlier projects concentrating upon developing repository infrastructure within Australia. The primary aim of CAIRSS is to provide a support service for repository managers as they face the challenges within the Australian university sector including ERA⁴ and ANDS. The presentation will look at the results of a CAIRSS survey completed by Australian university institutional research repository managers.

Methodology

In May 2010 CAIRSS conducted an online survey of Australian institutional research repository managers. The survey covered a number of areas including:

- Repository size and purpose
- Repository staffing
- Repository software
- Repository statistics
- Repository researcher IDs
- Links between repositories and institutional systems
- ANDS and repository research data
- Perceptions of repositories
- Feedback for CAIRSS

This conference paper focuses on the 'ANDS and repository research data' area of responses.

Survey questions were targeted to obtain a progress report of Australian research repositories, updating previously collected 2009 data. The online survey also included mechanisms for identifying upcoming challenges facing repository managers throughout 2010 and 2011. The online survey was followed by phone calls to individual repository managers where required for further data collation.

³ CAUL Australian Institutional Repository Support Service, 'CAUL Australian Institutional Repository Support Service (CAIRSS)', *CAIRSS*, 2009 <<http://cairss.caul.edu.au/www/index.htm>> [accessed 19 May 2010].

⁴ Australian Research Council, 'The Excellence in Research for Australia (ERA) Initiative - Australian Research Council (ARC)', 2009 <<http://www.arc.gov.au/era/default.htm>> [accessed 19 May 2010].

ANDS Background

One of ANDS' key drivers is the Australian Code for the Responsible Conduct of Research⁵ in which the management of research data is an important part. The code is not mandatory, but increasingly research funding is tied to code-conformance. ANDS has been provided with funding by the Federal Department of Industry, Innovation, Science and Research⁶ to be expended initially over a two year period to June 2011. This has now been extended to June 2013. Within its overall brief, ANDS seeks to respond to the Code by providing guidance for institutions seeking to comply by establishing effective management of research data. While many institutions already have their own policies for the conduct of research, the area of the management of research data, and in particular the exposing of such data in an appropriate manner, is relatively new. It is this within this area that ANDS seeks to provide assistance to researchers and institutions.

ANDS consists of various projects, two of which are most relevant to repository managers: (a) Seeding the Commons – establishing and seeding the Australian Research Data Commons which will support discovery and re-use of data through the use of Research Data Australia⁷ – and (b) the Metadata Stores project – enabling institutions to manage their local data registries in order to contribute to the Australian Research Data Commons. ANDS aims to investigate whether and how Australian university institutional repositories may be extended to cope with this new demand for datasets, or whether separate infrastructure is needed. With at least seven different types of institutional repository software in use across Australian universities, ANDS is also aiming to provide as much support and service as possible to enable this. Broadly speaking there are two approaches that organizations may take to implement metadata registries for research data: (a) they may do so by building on the institutional repository (most of which are managed by university libraries), or (b) may set-up separate systems, which are more likely to be managed by offices of research than by the library. The survey reported here and anecdotal evidence suggests that both models will be tested over the next few years.

Other projects being funded by ANDS include work on Metadata Hubs; Data Capture projects; Researcher Profile projects; and the Australian Research Data Commons Party Infrastructure Project.

Overview of the Australian Institutional Repository Community

The Australian university institutional repository community has grown over the past 8 years, to the point where 100% of the 39 universities now have institutional research repositories. Seven types of repository software are in use, with the majority of repositories managed on-site at institutions.

⁵ Australian Government, National Health and Medical Research Council, Australian Research Council, 'Australian Code for the Responsible Conduct of Research', 2007 <<http://www.nhmrc.gov.au/publications/synopses/r39syn.htm>> [accessed 26 May 2010].

⁶ Department of Innovation, Industry, Science and Research, 'Department of Innovation, Industry, Science and Research (DIISR)', 2010 <<http://www.innovation.gov.au/Pages/default.aspx>> [accessed 26 May 2010].

⁷ Australian National Data Service, 'Research Data Australia (RDA)', 2010 <<http://services.andis.org.au/home/orca/rda/>> [accessed 26 May 2010].

University institutional repositories within Australia are currently supported by CAIRSS⁸ (CAUL Australian Institutional Repository Support Service), having benefited from previous support and development projects since 2005 (such as the APSR⁹, ARROW¹⁰ and RUBRIC¹¹ Projects). Institutional repositories in Australia currently play a large role in the way universities report to the government for research reporting (Excellence in Research Australia, ERA¹²), and preparing repositories for ERA has been a key driver for many Australian repository managers throughout 2009 and 2010.

Due to the close relationship between the aims of ANDS and the Australian institutional repository community, the interaction between the two entities has been highlighted by the appointment of a specific role to liaise between the ANDS and the CAIRSS community.

Findings of surveys

With repository managers only just starting to look into their repository's capabilities in relation to research data, one of the aims of the survey was to find out what sort of awareness existed among the repository managers as to the tools and services offered by ANDS, including Publish My Data, Register My Data and Identify My Data. This was found to be higher than expected, with most of the repository managers stating that they have already looked at some of the services to see how they may relate to research repositories.

It was also found that the ERA exercise mentioned earlier has indeed been the focus of repository manager activity, with many people responding that the pressure placed upon them due to the ERA left little time for looking at ANDS, even though they were aware of its importance to them and their repository.

At this stage, more than half of repository managers who responded with a high awareness of ANDS and its aims, stated that their plans were still vague: just under half of the respondents are currently planning to use their institutional repository in its current form to house research data metadata; several stated that they do not plan to use their repository; and the rest of the respondents stated that they either don't yet know how their repository may or may not integrate into these activities, or have not yet made plans.

Awareness of the Australian Code of Responsible Conduct of Research exists in the repository community with 81% of respondents indicated they were aware of the code, with many discussing the ways they feel the code will impact on research repositories into the future.

As mentioned earlier, two models for data repositories are being examined and these shall be briefly outlined. At The University of Newcastle, library, research office and IT staff have developed a

⁸ CAUL Australian Institutional Repository Support Service.

⁹ Australian Partnership for Sustainable Repositories, 'Australian Partnership for Sustainable Repositories (APSR)', 2008 <<http://www.apsr.edu.au/>> [accessed 26 May 2010].

¹⁰ Australian Research Repositories Online to the World, 'Australian Research Repositories Online to the World (ARROW)', 2008 <<http://www.arrow.edu.au/>> [accessed 26 May 2010].

¹¹ University of Southern Queensland, 'Regional Universities Building Research Infrastructure Collaboratively (RUBRIC)', 2007 <<http://rubric.edu.au/>> [accessed 26 May 2010].

¹² Australian Research Council.

model where the metadata registry will be established as an adjunct to the existing institutional repository, with ‘data librarians’ curating metadata about research data. IT will supply storage services, which will be indexed by the library-based registry application and institutional triggers, such as changes in grant status will prompt data librarians to add or manage metadata. Metadata about research data will be stored in the same Fedora-commons-based repository as the current publications repository. Metadata records will be stored in an as-yet-undetermined format in XML in the repository, with URIs used to refer to other entities.

This contrasts with the approach being taken at the University of Melbourne where the research metadata store is a separate entity from the institutional repository. The registry is a RDF-based system, meaning that all metadata and relations will be expressed as relations between entities by URI.

Issues encountered from IRs and ANDS

Prior to the survey, questions had already arisen from some repository managers regarding appropriate use of metadata. Registry Interchange Format – Collections and Services RIF-CS¹³, a data interchange format based on ISO2146, created by the ANDS project for supporting the submission of metadata to a collections service registry, is raising questions for repository managers around Australia as to whether traditional institutional repositories are able to cope. With institutional repositories currently using bibliographic metadata, there isn’t yet a clear path as to the most appropriate data metadata format yet.

This presentation reports on surveys completed by Australian institutional repository managers highlighting some of the issues and concerns behind preparing repositories for this new role of data and dataset curation. Some of the specific issues discussed with the repository managers included looking at whether institutional policy exists behind curation and archiving of data and datasets and what an institution considers appropriate for inclusion in the ANDS framework. Anticipated interactions between traditional institutional repositories and ANDS will also be examined, specifically in relation to data formats both new and already in use. Lastly, the impact of this new curation role of institutional repositories will be examined in relation to some of the challenges already facing institutional repositories in Australian universities.

Learning from the past: Using Ontologies

One of the opportunities that is emerging as repositories begin to house data or be supplemented by data repositories, is to learn from some of the lessons of previous IR experiences. In particular there is now an opportunity to move to using a linked-data approach to metadata, using Berners-Lee’s 4 guidelines for linked data this means that each entity (Party, Activity, Service, in ISO-2146 terms) in a registry-entry is referred to using a URI. There are plans afoot to create end-points for entities in the data-space including referents for People and institutions via the Party Infrastructure Project and via authority services for grant-funders so that projects (activities) can be linked to a grant-ID. The problem facing early adopters is that this infrastructure is not yet available.

¹³ Australian National Data Service, ‘Registry Interchange Format - Collections and Services (RIF-CS)’, 2010 <<http://ands.org.au/resource/rif-cs.html>> [accessed 19 May 2010].

The Australian Digital Futures Institute¹⁴, and one of the authors is working with Newcastle University to establish their linked-data-ready data registry by creating a locally managed authority management service which will be populated with identifiers for Newcastle Staff and research projects – which can be synchronized with global identifiers and national services such as People Australia once that infrastructure has been created. Even when there are services that give URIs for research grants, institutions and potential all Australian researchers, there are still likely to be locally-relevant identifiers needed for things such as organizational units and self-funded research activities, so this name-authority service is likely to be needed long-term.

Applications of findings on a broader scale

The results of the survey presented in this paper could be of use to the broader institutional repository community in looking at the challenges facing repository managers who are looking at extending the traditional purpose of the institutional repository. Australia is approaching the curation, archiving and accessibility of data at a national level, but the approaches and resultant challenges may be applicable to any level of data curation. The issues that are being addressed, in particular those relating to the use of ontologies, the use of metadata schemas for research data, and the overarching management framework, could be of particular interest to those who are also facing the challenges of curating research data.

Bibliography

- Australian Government, National Health and Medical Research Council, Australian Research Council, 'Australian Code for the Responsible Conduct of Research', 2007 <<http://www.nhmrc.gov.au/publications/synopses/r39syn.htm>> [accessed 26 May 2010].
- Australian National Data Service, 'Australian National Data Service (ANDS)', 2010 <<http://ands.org.au/>> [accessed 19 May 2010].
- Australian National Data Service, 'Registry Interchange Format - Collections and Services (RIF-CS)', 2010 [accessed 19 May 2010].
- Australian National Data Service, 'Research Data Australia (RDA)', 2010 [accessed 26 May 2010].
- Australian Partnership for Sustainable Repositories, 'Australian Partnership for Sustainable Repositories (APSR)', 2008 <<http://www.apsr.edu.au/>> [accessed 26 May 2010].
- Australian Research Council, 'The Excellence in Research for Australia (ERA) Initiative - Australian Research Council (ARC)', 2009 <<http://www.arc.gov.au/era/default.htm>> [accessed 19 May 2010].
- Australian Research Repositories Online to the World, 'Australian Research Repositories Online to the World (ARROW)', 2008 <<http://www.arrow.edu.au/>> [accessed 26 May 2010].
- CAUL Australian Institutional Repository Support Service, 'CAUL Australian Institutional

¹⁴ University of Southern Queensland, 'Australian Digital Futures Institute (ADFI)', 2010 <<http://www.usq.edu.au/adfi/>> [accessed 31 May 2010].

Repository Support Service (CAIRSS)', *CAIRSS*, 2009
<<http://cairss.caul.edu.au/www/index.htm>> [accessed 19 May 2010].

Department of Innovation, Industry, Science and Research, 'Department of Innovation, Industry, Science and Research (DIISR)', 2010 <<http://www.innovation.gov.au/Pages/default.aspx>> [accessed 26 May 2010].

Kennan, M. A, and D. A Kingsley, 'The state of the nation: A snapshot of Australian institutional repositories', *First Monday*, 14 (2009).

University of Southern Queensland, 'Australian Digital Futures Institute (ADFI)', 2010
<<http://www.usq.edu.au/adfi/>> [accessed 31 May 2010].

University of Southern Queensland, 'Regional Universities Building Research Infrastructure Collaboratively (RUBRIC)', 2007 [accessed 26 May 2010].