

Publish My Data: A composition of services from ANDS and ARCS

Adrian Burton, Andrew Treloar
Australian National Data Service (ANDS)
{Canberra|Melbourne}, Australia
[\[adrian.burton|andrew.treloar\]@ands.org.au](mailto:{adrian.burton|andrew.treloar}@ands.org.au)

Abstract

Abstract – This paper first provides an overview of the e-Research environment in Australia, before introducing the two organizations (the Australian National Data Service – ANDS, and the Australian Research Collaboration Services - ARCS) whose services will be focus of the paper. Next the paper describes the problem that needs to be solved before talking about the requirements for management of data and metadata. The paper then discusses the main ARCS and ANDS services that will be combined to provide Publish My Data. The paper provides an overview of how this combined service will work for data that is already stored and data not yet stored, before concluding.

Keywords – data management; e-research; metadata

I. AUSTRALIAN E-RESEARCH BACKGROUND

A. Major Funding Programs

Australia has been investing in what we call e-Research (rather than trying to distinguish between e-Science, e-Humanities, e-Social Science, etc) since 2005. The major funding program is called the National Collaborative Research Infrastructure Strategy (NCRIS) [1]. This is providing A\$542M over the period 2005-2011 to fund up to 15 priority capabilities (mostly in the sciences).

A sixteenth capability, called Platforms for Collaboration (PFC), was funded to provide a range of infrastructure to support all researchers (not just those in NCRIS priority areas) [2]. PFC includes the National Computational Infrastructure (NCI) [3], the Australian Research Collaboration Service (ARCS) [4], and the Australian National Data Service (ANDS) [5]. ANDS allocation within PFC was \$A24M through to mid 2011. ARCS allocation was \$A26M for the same time period.

In the Federal Budget for 2009/10, both ANDS and ARCS received additional funding from the Education Investment Fund [6] under the Super Science program, an A\$1.1B boost for critical areas of scientific endeavour [7]. ANDS received an additional A\$48M over the remaining two years of ANDS to expand the Australian Research Data Commons (ARDC). ARCS received \$A97M over four years to significantly expand a national data fabric, as well as improve collaboration infrastructure.

B. ANDS overview

In late 2007, the then Department of Education, Science and Training (DEST) asked Monash University [8] as the lead agency to work with the Australian National University [9] and the CSIRO [10] on a project to establish the Australian National Data Service (ANDS). As a result of the successful achievement of these milestones, ANDS formally commenced operations in January 1, 2009.

ANDS is funded through FY 10/11 (Financial Years in Australia run from 1 July through 30 June) to make progress towards a number of ten-year objectives for data management:

A. A national data management environment exists in which Australia's research data reside in a cohesive network of research repositories within an Australian 'data commons'.

B. Australian researchers and research data managers are 'best of breed' in creating, managing, and sharing research data under well formed and maintained data management policies.

C. Significantly more Australian research data is routinely deposited into stable, accessible and sustainable data management and preservation environments.

D. Significantly more people have relevant expertise in data management across research communities and research managing institutions.

E. Researchers can find and access any relevant data in the Australian 'data commons'.

F. Australian researchers are able to discover, exchange, reuse and combine data from other researchers and other domains within their own research in new ways.

G. Australia is able to share data easily and seamlessly to support international and nationally distributed multidisciplinary research teams. ([11], p. 6)

To deliver against these objectives, ANDS had four NCRIS-funded inter-related programs of activity (Developing Frameworks, Providing Utilities, Seeding the Commons, Building Capabilities). ANDS also funded specific development activity towards the aims of the Providing Utilities and Seeding the Commons programs under the banner of the National e-Research Architecture Taskforce (NeAT).

ANDS has just completed a public consultation process on its plans to use the EIF money, as well as determining how to mesh this new expenditure with the existing NCRIS-funded business plan. The resulting Final Project Plan is available from the ANDS website. It introduces five new programs and

merges or adapts the existing NCRIS-funded programs. The resulting set of ANDS programs now comprises:

- Frameworks and Capabilities (merging of two of the original NCRIS-funded programs)
- Seeding the Commons (existing)
- ARDC Core Infrastructure (a rebadged and expanded Utilities program)
- Data Capture Infrastructure (new)
- Research Metadata Store Infrastructure (new)
- Public Data Access Infrastructure (new)
- ARDC Applications Infrastructure (new)

The work described in this paper is being undertaken as part of the ARDC Core program.

C. ARCS overview

The ARCS mission is to enable and enhance research through the provision of long-term eResearch support, services and tools. The use of ARCS Services and Tools are free of charge to the Australian research community unless specified otherwise. ARCS provides services across a range of categories.

Under Authorisation Services, ARCS is working with eResearch Service Providers to build a unified authorisation infrastructure for eResearch Services and Tools, and with national discipline-based research groups to provide tools, services and support for authorisation and authentication issues.

Under Collaboration services, ARCS provides a wide range of Collaboration Services and Tools allowing researchers, groups, and research communities to share ideas and collaborate across organisational boundaries. These include video collaboration, web-based collaboration and custom solutions. ANDS is a heavy user of the ARCS EVO video collaboration solution [12] and Plone [13] web-based collaboration solution to manage its distributed activities.

Under Systems services, ARCS provides assistance and expertise to Australian researchers who submit compute jobs to one or more of the High Performance Computing (HPC) facilities available through the ARCS Grid, supplementing the NCI. The ARCS Grid allows users to collaborate with remote groups and to conduct their research on the facilities that best suit their needs, regardless of its location.

Under Data services, ARCS provides researchers with tools and services to

- allow research groups and individuals to store, maintain and share their data
- mirror large datasets between two or more data centres,
- make databases accessible by research collaborators based in different organisations,
- receive and store large data sets from international partners,
- collect datasets locally and make them accessible nationally,
- provide storage targets for data generated with scientific instruments.

II. PROBLEM WE ARE TRYING TO SOLVE

Previously there has been an expectation that scholars will make their findings public, but not necessarily the data, observations, models, recordings, etc upon which those findings are based. A number of changes in society and technology have combined to create an increased expectation that scholars will make their data public. A few scattered examples might include the following:

- High profile cases of scientific fraud have put pressure on researchers and publishers to provide ways of verifying claims both before and after publication [14].
- Debate on science-based public policy naturally questions the data upon which it is based [15].
- In an attempt to stimulate innovation and avoid duplication, funding agencies around the world are introducing guidelines, suggestions and mandates for making public the data generated with their funds [16].
- New directions in scholarly communications are also moving towards increased online linking of data and publications.

Many research disciplines have seen the strategic opportunities in aggregating data into formal data archives or databanks; many are trying to establish distributed data networks. Some disciplines, such as astronomy and particle physics have a small number of researchers (or large instruments) that produce very large data volumes and whose data needs are generally well catered for. But there is also a long tail [35] of scientists and humanist researchers who are much greater in number but who each produce much smaller data volumes. This group has less in the way of infrastructure and services to enable them to respond to these new expectations.

ANDS and ARCS are developing services and infrastructure to enable researchers to “publish” their data. Research and analysis work in ANDS has identified a number of basic service functions that in most cases are required to enable the sharing and re-use of data through its publication. ANDS refers to these as Data Sharing Verbs. A summary of these verbs (more detail is available at [17]) is:

- **Create:** the process of creating data outputs from research, with an emphasis on automated collection of metadata about the data outputs as early in the research process as possible.
- **Store:** the storage of the data outputs on stable, well-supported infrastructure.
- **Identify:** assigning a persistent identifier to the data output to enable citation of the data as a first-class research object and to provide an indirection layer between the citation and the storage location to facilitate management of the object.
- **Describe:** associating metadata with the data output that will assist with its storage, preservation, discovery, access and exploitation.

- **Register:** registering some or all of this metadata in a location to enable discovery and re-use.
- **Discover:** using a variety of tools and services to discover the existence of a research data output and assess its relevance or quality.
- **Access:** gaining access to the actual research data output, which might involve an authentication and authorisation step.
- **Exploit:** making use of the available metadata to exploit and re-use the data, including such things as data fusion and visualisation.

ANDS and ARCS are providing a number of services to support these individual functionalities and enable researchers or their organisations to publish their data. In a loosely coupled approach, these functionalities can be sourced from many places and brought together in many combinations to produce the same effective result. This paper describes some ANDS and ARCS services that enable some of these functionalities. It also describes a few possible orchestrations of these services to achieve the publication of data in support of sharing and re-use.

III. METADATA AND DATA ISSUES FOR DATA PUBLICATION

A. Metadata storage options

Publication of data alone is not enough. In order for the data to be discoverable and re-used, it is necessary to provide discovery information to enable a user to discover the existence of a data collection, assessment information to enable a user to decide if the discovered data is of interest to them, and re-use information to enable a user to make use of the data once they have decided is of interest to them.

This data needs to be captured as close to the time of creation as possible [18] and managed somewhere. Most data stores are not equipped to manage this level of metadata at a collection level. ANDS has an active program of work to investigate ways of using existing institutional repositories as metadata stores. ANDS will also be funding the creation of add-ons to existing data store infrastructure as well as a standalone solution.

B. Data storage options

One of the notable characteristics of the Australian e-Research environment described above is that ANDS has not been funded under NCRIS to provide data storage. The amount of money available was inadequate, and under the Australian Code for the Responsible Conduct of Research [35] this is viewed as an institutional responsibility. The recent injection of EIF funds has not changed this picture, as ARCS has been explicitly charged with building out a national data fabric. As a result, ANDS needs to rely on a range of other locations where data might be stored.

The first is in ‘traditional’ institutional repositories. Through the ASHER program [19], all

Australian universities have implemented one of these. Although these have been designed for document objects, many of them can also be used for data [20]. As an example, the ARROW [21] repository at Monash University contains both protein crystallography raw image data [22] and ethno-musicology fieldwork recordings [23].

The second is in institutional data stores. A number of Australian universities and large research institutions are putting place specialized data stores, optimized for large numbers of large objects. An example is the Monash University Large Research Data Store (LaRDS) [24]. These are often based on very different technologies to institutional repositories, such as SRB [25], iRODS [26], or NetCDF/OpeneNDAP [27].

The third location is on the ARCS data fabric.

IV. RELEVANT ARCS AND ANDS SERVICES

A. ARCS Data Fabric

ARCS has been funded to provide a range of data tools and services. ARCS is a federated organization, built on top of State and Territory e-Research service providers, but researchers want to collaborate across these boundaries. ARCS is therefore implementing (among other things) a national data fabric to provide a production service allowing researchers, research groups, research organizations and research communities to store their data. The current version of the data fabric provides free storage of 25GB for each researcher, with additional storage provided at near hardware cost for three years. Additional managed services such as enduring storage, archiving, backups, and data replication can also be provided at the effective cost to ARCS. The main technology for the ARCS data fabric is iRODS, and ARCS is building both web-based and desktop interfaces. More details about the ARCS data fabric are available online [28]. It should be noted that the size, functionality, and range of services available through the data fabric will increase substantially over the next four years as a result of the additional funding received. The details of precisely what will happen are currently (October 2009) being worked out and will need to be subject to a consultation process prior to implementation commencing in 2010.

B. ANDS Identify My Data

The ANDS Identify My Data service enables researchers and their organisations to allocate persistent identifiers to datasets and other elements of the network of information. Persistent identifiers can be used to refer to a dataset, without the danger of the reference becoming unresolvable when the dataset changes URL. This persistence is a prerequisite for datasets to participate in the long-term scholarly communications cycle as ‘published’ items. Once included in the scholarly record, the reference must persist for a reasonable amount of time, despite underlying changes the ownership and organisation of data stores.

The service is presently powered by an “in the cloud” Local Handle Server open to all organisations and individuals in the Australian research and innovation sector. A standard usage scenario would involve a data facility at an Australian research facility becoming a registered user of the service. The data facility makes web service calls to the ANDS service when an item is deposited. The ANDS service responds by minting a globally unique identifier. The data facility allocates the identifier to the dataset and updates the location information with the ANDS service whenever its location changes.

The ANDS service uses the global Handle System [29] infrastructure to ensure resolution of the identifier. This identifier is in the form of a numerical prefix and an alphanumerical suffix; for instance 1959.1/5863. The Handle System has a design that supports uniqueness at each level: the global infrastructure ensures that it only assigns the prefix (1959.1 above) to a single organisation or name authority. The software run at each name authority ensures that a suffix is only assigned once and never re-used. The Handle System also provides a resolution service to ensure that users can resolve from a handle to the corresponding URL, and there is also software to allow the binding between a handle and a URL to be maintained as the location of the identified object changes. It should be noted that the Handle System only provides the infrastructure to ensure uniqueness – it is up to individuals and organisational processes to ensure that the binding between handle and location is maintained.

ANDS provides the service and infrastructure to enable research organisations to persistently identify their data. The persistent identifier service can also be used to provide a more resilient way to identify and refer to other critical elements of the networked environment such as web services, harvesters, harvester data providers, semantic web assertions, annotations, registries etc. Each of these can be assigned a handle as a form of future-proofing.

C. ANDS Register My Data

The ANDS Register My Data [30] services enable researchers and research organisations to register public descriptions of data collections. These descriptions include information about the collection, any parties involved with it, any research activities related to it, and any services that can be used to access it. These entities are taken from the draft ISO2146 framework [31]. The collection description also includes reference to the technical metadata that would facilitate re-use, where this is available. The registered descriptions are published in various online browse and discovery environments in ANDS and beyond ANDS, nationally and internationally. The publication process makes the data collection descriptions indexable by the big internet search engines.

The Register my Data Service is currently powered by a registry and harvest application (ORCA [32]) with the first point of publication being Research Data Australia [33]. Under a typical usage

scenario, an organisation’s data repository registers with the service and the registry harvests the collection descriptions information at regular intervals from the repository [34].

V. PUBLISH MY DATA

Publish My Data is a line of ANDS services that (when combined with other non-ANDS services) enable organisations and individuals to make their data public. The publication of data is a multi-party multi-layered course of action that involves in most cases a combination of the ANDS Data Sharing Verbs.

ANDS and ARCS provide some infrastructure and services to support some of these functions (eg store, identify, register, discover). Research organisations and other service providers also provide some of these same service functions. Some applications such as institutional repositories combine a number of these functions into the one environment. A mix and match approach is clearly possible and desirable from the authors’ perspective. This paper describes three possible ways that users might interact with a Publish My Data service built on top of the ANDS Data Sharing Verbs.

A. Distributed Publish My Data

Because Publish My Data is a set of orchestrations of a number of these service components to achieve the publication of data, the component functionalities can be embedded in institutional data facilities so that the end user (researcher) would not know that ANDS was even involved. This scenario would run something like this:

1. A *researcher* **creates** a dataset of observations on a laptop in the field
2. The researcher **stores** the dataset in some institutionally-supported datastore
3. On deposit, the *archive* **identifies** the dataset with a persistent identifier provided by ANDS via web service
4. On deposit, the *researcher* **describes** the data collection (assisted by auto metadata provision).
5. The *archive* **registers** the collection description with the ANDS registry (among others) using automated harvester
6. End users **discover** the collection description through any number of browse, discovery and or search portals (including discipline specific ones)
7. End users **access** the data in the institutional data store
8. End users use the related technical metadata (calibrations, scales, classification schemes, vocabularies, etc) to **exploit** the data

This generic process is shown in Figure 1. The details of the user-interface have not been shown. This is because the calls to the underlying services will be made via software that is written by the institution at which the researcher is based. ANDS just provides the functionality – how it is orchestrated and presented is (in this model) out of

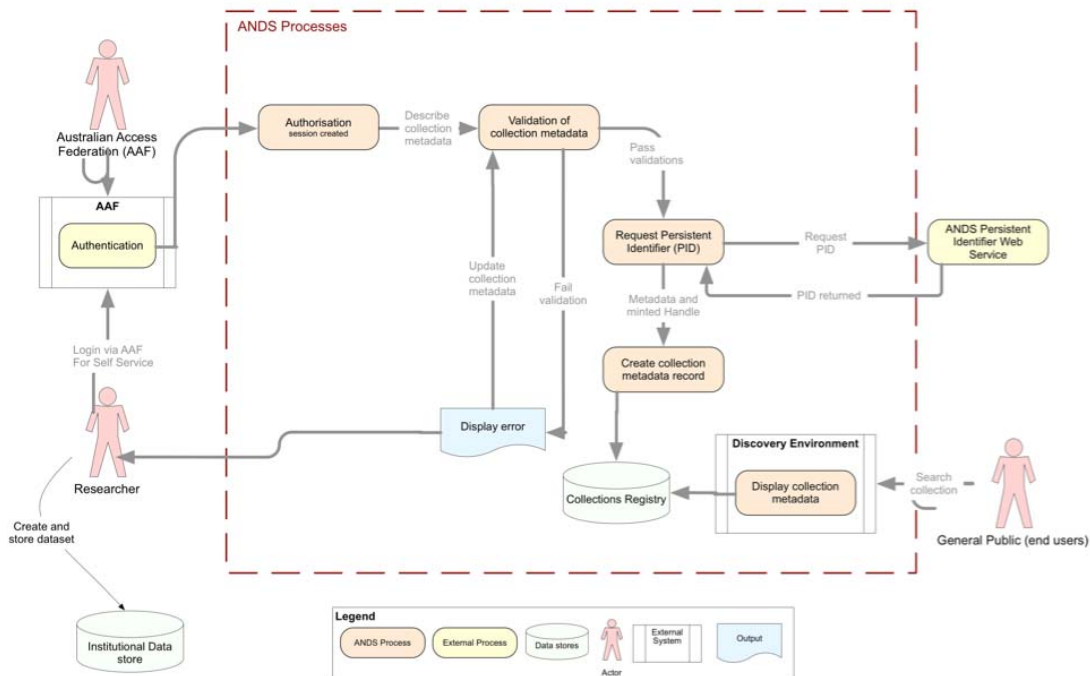


Figure 1: Distributed Publish My Data

ANDS’ control. The business logic and workflow control would be embedded in the institutional software.

From ANDS point of view this is a distributed model of Publish My Data because the business of publication is actually taking place at each data archive (rather than centrally at an ANDS website). ANDS is simply providing some enabling services to data archives. This is the standard model for ANDS operations.

In the next two sections, two other “Publish My Data” scenarios are described which are relatively more centralised, because more of these activities are done at the ANDS website. These scenarios also involve the ARCS Data Fabric providing the **store** functionality. The ARCS Data Fabric **store** and **access** services are coupled with the ANDS **identify**, **describe**, **register**, **discover** services to create the desired outcome of published data. These scenarios differ from the one described already in that there is no institutionally-supported datastore. It is not unheard of for there to be no internet accessible storage infrastructure readily available for the storage of valuable research data sets. In these cases the ARCS Data Fabric can provide that functionality to the Australian research and innovation community. “Published data” needs to be reliably stored and network accessible.

However, the ARCS Data Fabric is not a collection management environment; it is a storage and access environment. Therefore in these scenarios there is no archive manager. From an ANDS perspective these are more centralised Publish My Data workflows because the functionality is no longer embedded with data archives; rather the

researchers come to an ANDS Publish My Data website to **identify**, **describe**, and **register** their own data.

B. Centralised Publish My Data (Data starts in the ARCS Data Fabric)

Work has begun to make the ANDS **identify**, **describe**, **register**, and **discover** functionalities available to data stored in the ARCS Data Fabric. This comprises a loose coupling of the services provided by both organisations to value-add in both directions. The first workflow assumes that data is already stored in the ARCS Data Fabric. The ARCS Data Fabric interface contains a Publish My Data button which calls an ANDS web page giving access to these ANDS services. Once again, the details of the user-interface are not described here. The workflow would go something like this:

1. A researcher **creates** a dataset of observations on a laptop in the field
2. The ARCS Data Fabric **stores** the dataset
3. When the researcher requests that this data collection be made public, the researcher **identifies** the dataset (ie, allocates it a persistent identifier provided by ANDS via the web page)
4. On the same ANDS web page the researcher **describes** the data collection (assisted by auto metadata provision).
5. The submission button on that web page **registers** the collection description with the ANDS registry.
6. End users **discover** the collection description through a number of browse, discovery and or search portals

7. End users **access** the data in the Data Fabric
8. End users use the related technical metadata (calibrations, scales, classification schemes, vocabularies, etc) to **exploit** the data

C. Centralised Publish My Data (Data starts on the desktop)

A not uncommon scenario is that valuable research data is held on the desktop, or sometimes on a thumb drive in the bottom drawer. This scenario differs from the previous in that a researcher visits the ANDS website as a first port of call and uploads the data to the ARCS Data Fabric as part of the Publish My Data workflow. The workflow would be something along these lines:

1. A *researcher* **creates** a dataset of observations on a laptop in the field
2. The researcher visits the ANDS web site and elects to “Publish My Data”. The data set is uploaded via the web page, and the *ARCS Data Fabric* **stores** the dataset
3. On the same web page the *researcher* **identifies** the dataset (ie, allocates it a persistent identifier provided by ANDS via the web page)
4. On the same ANDS web page the *researcher* **describes** the data collection (assisted by auto metadata provision).
5. The submission button on that web page **registers** the collection description with the ANDS registry.
6. **discover** as in previous
7. **access** as in previous
8. **exploit** as in previous

D. Workflow considerations

The use of the word “workflow” in the preceding discussion should not be interpreted to mean that the implementation would necessarily involve a formal workflow management system. The initial implementations of Publish My Data are encoding the service calls and error-handling within the business logic layer of conventional web applications.

As the underlying services have standards-based APIs or are accessible via web-services (either REST or SOAP) it would be possible to create novel service compositions using a workflow engine, but this has not yet occurred.

VI. INTERNATIONAL COMPARATORS

A. National approaches

So, how does this approach compare with equivalent activities in other countries? Australia is the only country to date that is taking a whole-of-country, whole-of-research approach to research data management.

The United States has announced its Sustainable Digital Data Preservation and Access Network Partner (DataNet) program [36] and has made two calls for proposals. The first set of funded proposals has now been announced. These are focused on

institutional consortia, with a specific discipline focus.

The United Kingdom undertook an investigation into a UK Research Data Store [38] in 2008. This undertook a feasibility study and received further funding for a limited pathfinder implementation. The UK Joint Information Systems Committee (JISC) also issued a funding call in 2009 for Data Management Infrastructure, seeking to fund 6-8 pilot projects at up to £250K each.

B. Data deposit interfaces

A number of organizations in the UK and Europe have sought to provide data deposit interfaces for specific disciplines as part of their remit. How do these compare to Publish My Data?

1) UKDA-store

The UKDA-store [40] is provided by the UK Data Archive [41] as a repository for managing and sharing research outputs, with the focus being on data outputs. The UKDA-store aims to support the sharing of data and other research outputs (primarily across the social sciences and humanities) without offering a formal archive. Registered researchers whose research has been funded by the UK Economic and Social Research Council can:

- “upload a range of digital objects to the repository including statistical data, databases, word documents, audio-visual materials
- assign permissions to individuals and/or groups to enable access to the materials in the repository
- ask for data to be formally lodged with the UK Data Archive” [40]

2) DANS EASY

The EASY system [42] is provided by the Dutch Data Archiving and Networked Services [43]. EASY is intended for researchers in the social sciences (social science, behavioural science, socio-cultural sciences) life sciences and medicine and geospatial sciences, as well as some of the humanities. EASY stores deposited data in a “permanent and sustainable manner, according to the guidelines of the international Data Seal of Approval. The data are made available to other researchers under specific conditions in accord with the depositor.” [42].

3) Discussion

UKDA-Store and DANS EASY are different to the ANDS Publish My Data (PMD) service on a number of important dimensions.

Firstly, they are both constrained to a subset of disciplines, reflecting their intended user-base or funding source. ANDS intends to support all research disciplines.

Secondly, they are essentially institutional repositories with a wider remit. Both are either based on Fedora (UKDA-Store) or being re-written on top of Fedora (DANS-EASY) and provide fairly standard repository deposit interfaces, complete with the capture of DC metadata. ANDS PMD is not an institutional repository – it is a set of service compositions across more than one organization (local institutions, ARCS and ands)

Thirdly, they are single integrated offerings. ANDS PMD is designed to be delivered in a number of different flavours on top of different underlying storage architectures. ANDS PMD is also designed to allow individual institutions to repurpose parts of it for their own ends or embed it within their own research output management systems.

VII. CONCLUSION

We believe that this decomposed approach is not novel in conception, but novel as a national supported implementation. Identifying the key component functions involved in making data public has informed a highly flexible approach to providing support services and infrastructure. This service-oriented approach to data publication allows a variety of service providers to be orchestrated in ways appropriate to each user's context. The strength of the approach is the way that it allows the natural heterogeneity of real systems and different organisations to flourish, enabling diversity without attempting to replace it.

VIII. REFERENCES

- [1] <http://ncris.innovation.gov.au/Pages/default.aspx>
- [2] <http://pfc.org.au/>
- [3] <http://www.apac.edu.au/nci/>
- [4] <http://arcs.org.au/>
- [5] <http://ands.org.au/>
- [6] <http://www.deewr.gov.au/HigherEducation/EIF/>
- [7] <http://innovation.gov.au/Section/AboutDIISR/FactSheets/Pages/SuperScienceInitiativeFactSheet.aspx>
- [8] <http://www.monash.edu.au/>
- [9] <http://www.anu.edu.au/>
- [10] <http://csiro.au/>
- [11] Australian National Data Service Technical Working Group (ANDS TWG) (2007). Towards the Australian Research Data Commons. Retrieved July 24, 2008, from <http://www.pfc.org.au/bin/view/Main/Data>.
- [12] <http://evo.arcs.org.au/>
- [13] <http://plone.org/>
- [14] Maris, E. "Should journals police scientific fraud?", *Nature* 439, 520-521 (2 February 2006), doi:10.1038/439520a; Published online 1 February 2006
- [15] As an example see <http://www.rightsidenews.com/200901223408/energy-and-environment/scientists-data-challenge-new-antarctic-warming-study.html>
- [16] For example in Australia <http://www.nhmrc.gov.au/publications/synopses/r39syn.htm> and in the USA <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-05-022.html>
- [17] Burton, A. and Treloar, B., "Designing for Discovery and Re-Use: the ANDS Data Sharing Verbs Approach to Service Decomposition", Proceedings of the International Data Curation Conference (IDCC) 5, London, December 2009 (in press).
- [18] Treloar, A. and Wilkinson, R. "Rethinking Metadata Creation and Management in a Data-Driven Research World". Proceedings of IEEE e-Science 2008, December, Indianapolis.
- [19] <http://www.innovation.gov.au/Section/AboutDIISR/FactSheets/Pages/AustralianSchemeForHigherEducationRepositories%28AHSER%29FactSheet.aspx>
- [20] Treloar, A. and Harboe-Ree, C. "Data management and the curation continuum: how the Monash experience is informing repository relationships". Proceedings of VALA 2008, Melbourne, February.
- [21] <http://arrow.edu.au/>
- [22] <http://preview.tinyurl.com/nl88yg>
- [23] <http://preview.tinyurl.com/kv9692>
- [24] <http://www.monash.edu.au/eresearch/activities/lards.html>
- [25] http://www.sdsc.edu/srb/index.php/Main_Page
- [26] <https://www.irods.org/>
- [27] <http://opendap.org/>
- [28] <http://www.arcs.org.au/products-services/data-services/arcs-data-fabric-1/arcs-data-fabric>
- [29] <http://www.handle.net/>
- [30] <http://ands.org.au/services/register-my-data.html>
- [31] <http://www.nla.gov.au/wgroups/ISO2146/>
- [32] <http://www.globalregistries.org/orca.html>
- [33] Beta version at <http://services.ands.org.au/pages/>
- [34] Using the OAI-PMH protocol with RIF-CS (see <http://ands.org.au/resource/techdocs.html>)
- [35] National Health and Medical Research Council/Australian Research Council/Universities Australia. The Australian Code for the Responsible Conduct of Research. October 2007. Retrieved November 4 from <http://www.nhmrc.gov.au/publications/synopses/r39syn.htm>
- [36] Anderson, Chris. "The Long Tail". *Wired*, October 2004. <http://www.wired.com/wired/archive/12.10/tail.html>
- [37] http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503141
- [38] <http://ukrds.ac.uk/>
- [39] http://www.jisc.ac.uk/fundingopportunities/funding_calls/2009/05/grant0709.aspx
- [40] <http://store.data-archive.ac.uk/store/>
- [41] <http://www.data-archive.ac.uk/>
- [42] <http://easy.dans.knaw.nl/>
- [43] <http://www.dans.knaw.nl/>