Designing for Discovery and Re-Use: the 'ANDS Data Sharing Verbs' Approach to Service Decomposition

Adrian Burton, Andrew Treloar,

Deputy Directors,

Australian National Data Service

Abstract

This paper describes how the Australian National Data Services (ANDS) is designing systems to support data sharing and Re-use. The paper commences with an overview of the setting for ANDS, before introducing ANDS itself. The paper then structures its discussion of ANDS services for Re-use in terms of the ANDS Data Sharing Verbs: Create, Store, Describe, Identify, Register, Discover, Access and Exploit. For each of the data verbs, a rationale for its importance is provided together with a description of how it is being implemented by ANDS. The paper concludes by arguing for the data verbs approach as a useful way to design and structure flexible services in a heterogenous environment.[1]

The Australian National Data Service

Origins

In late 2007, the then Department of Education, Science and Training (DEST) asked Monash University as the lead agency to work with the Australian National University and the Commonwealth Scientific and Industrial Research Organisation (CSIRO) on a project to establish the Australian National Data Service (ANDS). ANDS is formally part of the Platforms for Collaboration[2] capability within the National Collaborative Research Infrastructure Strategy (NCRIS)[3]. Other components of Platforms for Collaboration include the Australian Research Collaboration Service (ARCS)[4] and the National Computational Infrastructure Program[5]. Most of 2008 was spent on the establishment phase, prior to the Australian National Data Service[6] (ANDS) formally commencing operations in January 1, 2009.

ANDS is funded through FY 10/11 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, re-use 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. (Australian National Data Service Technical Working Group [ANDS TWG] (2007), p. 6)).

Funding

ANDS was originally funded for a total of A\$24M for the period from mid 2007 (the start of preparatory phase) through to mid 2011 (the end of the NCRIS funding program). In the Federal Budget for 2009/10, ANDS received an additional A\$48M over the remaining two years of ANDS from the Education Investment Fund[7]. ANDS responded to the new funding by significantly reworking its project plan. The results of this work were accepted by the funding agency in October 2009 and the Final Project

---

[2] Platforms for Collaboration: http://pfc.org.au/
[3] National Collaborative Research Infrastructure Strategy http://ncris.innovation.gov.au/
[4] Australian Research Collaboration Service http://arcs.org.au/
[5] National Computational Infrastructure Program http://www.apac.edu.au/nci/
[6] Australian National Data Service http://ands.org.au/
[7] Education Investment Fund http://www.deewr.gov.au/HigherEducation/EIF/

Plan is available online[8]. This plan includes a significantly reworked set of seven co-ordinated programs: Frameworks & Capabilities, Data Capture, Research Metadata Stores, Seeding the Commons, ARDC Core, Public Data Access, and Applications.

Vision

The vision for ANDS is usually summarized as "More researchers re-using data more often"[9]. This requires activity across a whole range of areas from data management policies and plans through to discovery services. The purpose of this paper is to focus on some of the implications of this vision, and how they are playing out in the realization of some of our services.

Implications

ANDS is focused on data that will be made available for re-use. This doesn't have to be data that is publicly available, but it does need to be shareable. The design of ANDS' discovery services is informed by the need to facilitate re-use both within disciplines (which in some cases may already be well served) and across disciplines. This latter use case is of greater importance, as ANDS is charged with facilitating cross-disciplinary activity as a way of boosting Australia's research performance.

In order to facilitate re-use, it is useful to help people discover the data in context. The design of the initial release of our discovery services has been informed by feedback that a traditional meta-driven portal would not be sufficient. Instead, ANDS is working with the ISO 2146[10] draft standard, which is based around four different first class entities: Collections, Parties, Activities, and Services. ANDS harvests descriptions of these entities from a range of sources into our Collections Registry[11]. and builds human-readable and machine-spiderable webpages for each entity instance, as well as building connections between them. The results of this process can be viewed on the ANDS Research Data Australia pages[12]. Over time, we will be adding more descriptions of Activities and Services, and increasing the richness of the pages.

Because the pages are machine-spiderable, we anticipate that most of our access will come in through searches on Google and other web search engines. Users will then be able to follow links from a data collection to the project that produced it or the researcher responsible. This provides valuable context to enable a user to decide if they want to take the next step and access the data for potential re-use.

ANDS has chosen to base its discovery services on data collections, not data items. This is because collections (if correctly defined[10)] are the right level for effective assessment of re-usability. This decision also ensures that the collections registry doesn't become filled with a large number of identical small data sets (for instance, one per day from a particular environmental sampling location).

---

[8] Australian Research Data Commons Final Project Plan http://ands.org.au/infrastructure.html
[9] With apologies to Bicycle Victoria http://www.bv.com.au/
[10] ISO 2146 Project http://www.nla.gov.au/wgroups/ISO2146/
[11] ANDS Collections Registry https://services.ands.org.au/home/orca/search.php
[12] Research Data Australia http://services.ands.org.au/home/orca/rda/index.php

ANDS Data Sharing Verbs

Genesis

Given this context, how has ANDS been thinking about the things that need to happen to realize this vision and the resulting high-level service architecture? The point at which the thoughts of the two authors crystallized was during a discussion led by Steve Androulakis from Monash University about his plans to rethink the system design for the TARDIS[13] -federated diffraction image publication repository (Androulakis, 2008). We realized that we could generalize his approach and use it both to describe and design the various components of ANDS and the systems it interacts with. The advantage of thinking about elements in this way is that it decouples the operations from the systems that perform them. This is critical for ANDS because we have little control or even influence over some of the infrastructure pieces on which we rely.

This paper describes the result of putting that insight into practice. As we have worked through this process, we have come to refer to this way of looking at ANDS as 'Data Sharing Verbs'[14]. The current list of Data Verbs contains Create, Store, Identify, Describe, Register, Discover, Access, and Exploit. The verbs are not meant to cover all functions related to research data. As outlined earlier, ANDS is currently focusing on the re-use and sharing of data; we may add additional verbs (such as Define or Preserve) over time.

The verbs are not an end in themselves, but are a way to identify key functions required to support the re-use of data and (this is the real point) to help map to each function a variety of systems, services and organizations that might provide that functionality in a given context. For each of the verbs below, the heading shows which of the ANDS long-term objectives it supports.

Create (Supports ANDS Objective B)

Create in this context should be taken to include 'collect' (for disciplines with an observational focus, including humanities, sciences and social sciences). The ICT revolution has totally transformed the amount of digital data being created, and this is the source of the currently perceived data deluge. For example, individual researchers continue to record their observations, but now so too do powerful satellites. Create is a fairly self-evident function, because some agent must create data at some point of time, if not there is nothing to share. And as noted there is no need for any encouragement to create more data. ANDS does not provide support for the creation of data objects, but it does care how this takes place. In particular, ANDS believes that the creation of metadata about data objects can take place most cost-effectively at the earliest point possible in the data lifecycle (Treloar & Wilkinson, 2008). A part of ANDS' engagement with researchers and institutions, therefore, is to focus on talking through the kinds of metadata that could be collected early in the process, and how these might be augmented to support later re-use. As part of a newly-funded program of work we will be commissioning the development of solutions for data capture from research instruments consistent with these aims.

---

[13] Targeting Academic Research for Deposit and Disclosure http://tardis.edu.au/
[14] It may, of course, not be entirely coincidental that both authors have a background in linguistics (Treloar in Germanic languages, Burton in Indic).

Store (Supports Objectives A, C)

Within a vision for a data commons, the need for stable, web-accessible storage is fundamental. One cannot share, discover, curate, or re-use data that has not been retained somewhere. The engineering challenges of retaining a petabyte of data for a hundred years are non-trivial (Jantz & Giarlo, 2005; Rosenthal, 2007). There are significant policy questions that need to be addressed for many of the stakeholders around the benefits, responsibilities, and optimal arrangements for the storage of data.

As described above, ANDS is not funded to provide storage, but ANDS does care that data is stored appropriately (appropriately in this context means by someone who cares and where the data is likely to persist for a reasonable period of time). ANDS does not yet require something like Trusted Digital Repository certification for the data stores, but is looking at instruments like the Dutch Data Seal of Approval[15] as a possible approach. In practice, there are three locations where ANDS believes most data will be stored.

The first is in 'traditional' institutional repositories. Although these have been designed for document objects, many of them can also be used for data (Treloar & Harboe-Ree, 2008). As an example, the ARROW[16] repository at Monash University contains both protein crystallography raw image data[17] and ethno-musicology fieldwork recordings[18].

The second is in institutional data stores. A number of Australian universities and large research institutions are putting in place specialized data stores, optimized for large numbers of large objects. An example is the Monash University Large Research Data Store (LaRDS)[19]. These data stores are often based on underlying technologies, protocols and file formats (such as SRB[20], iRODS[21] or NetCDF/OpeNDAP[22]) that are very different to those used for conventional institutional document repositories.

The third location is on a national data fabric, such as that being built by the Australian Research Collaboration Service[23] (ARCS funded under the same program as ANDS). As part of its work, ARCS is developing and providing a number of tools that allow researchers and research groups to store, share and transport data across institutional boundaries. The main service through which these capabilities are delivered is the ARCS Data Fabric[24]. This is available to all Australian researchers and can be accessed through a variety of interfaces including web browsers and operating system integration.

ANDS plans to work with all three categories of solutions to have our concerns about availability, persistence, and an appropriate level of metadata be met in the services provided.

---

[15] Data Seal of Approval http://www.dans.knaw.nl/en/data_deponeren/dans_keurmerk/
[16] Australian Research Repositories Online to the World http://arrow.edu.au/
[17] Making long URLs usable http://preview.tinyurl.com/nl88yg
[18] Making long URLs usable http://preview.tinyurl.com/kv9692
[19] LaRDS overview http://www.monash.edu.au/eresearch/activities/lards.html
[20] Storage Resource Broker http://www.sdsc.edu/srb/index.php/Main_Page
[21] Integrated Rule-Oriented Data System https://www.irods.org/
[22] Open-Source Project for a Network Data Access Protocol http://opendap.org/
[23] Australian Research Collaboration Service http://arcs.org.au/
[24] ARCS Data Fabric start page
http://www.arcs.org.au/products-services/data-services/arcs-data-fabric-start-here

Describe (Supports Objectives B, E, F)

The more information there is about data, the greater the value of the data. Contextual information enables storage, preservation, discovery, access and exploitation of research data. Unfortunately, the cost involved in creating that added value is significant – metadata are expensive and difficult to obtain. At its broadest the "describe" function includes any information that will assist storage, preservation, discovery, access and exploitation of research data. This is broader than most conceptions of metadata. The aggregation of the dataset and all these kinds of information (wherever they might be) is a more powerful conception of a data collection. There is good scope for using protocols such as the OAI-ORE for framing collections in these terms.

To support the goal of "Discovery within Context" (see above), ANDS seeks descriptions of collections that include descriptions of the people, organizations, and activities that gave rise to the data. The result is a browsable mesh of information about researchers, research organizations, research projects and their data. The framework that informs this approach is (draft) ISO 2146, Registry Services for Libraries and Related Organisations.

ANDS has defined an XML schema for these descriptions based on ISO 2146. This format is called the Registry Interchange Format – Collections and Services (RIF-CS).[25]. RIF-CS provides a lot of flexibility for transferring between registries any information about data collections, parties, activities, services. As an interchange format it provides a vehicle for transferring structured information between registries without stipulating particular structures or classification schemes. Within the context of ANDS operations, the creation of these descriptions in RIF-CS format can occur in a number of ways.

The most straightforward is for the descriptions to be created manually by the owner (or someone with the right authority). Of course, this requires a degree of familiarity with XML and an appropriate editor. ANDS is currently building software to assist with this.

Another option is for the RIF-CS to be generated automatically in software. For this to occur, there need to exist sufficient metadata at the level of the collection being described. This approach has been used successfully in pilot with the Australian Social Science Data Archive, the Australian National University Supercomputing Facility, several Institutional Repositories, and the University of Melbourne e-Scholarship Centre.

A third option is for ANDS to harvest well-formed XML collection descriptions in a different format and transform them on ingest into the ANDS Collections Registry. This approach has been used successfully with some of the marine collections available through Research Data Australia. In this case, ISO 19115/19139 metadata conforming to the Australian Marine Community Profile (MCP) (BlueNet version 1.4) was harvested and transformed.

---

[25] RIF-CS is extensively documented at the Global Registries website http://www.globalregistries.org/rifcs.html

Identify (Supports Objectives E, F)

The Identify step can happen either before or after the Describe step. In other words, this isn't a strictly linear sequence of operations. Identify involves assigning a persistent identifier of some sort to the data collection. This provides at least two advantages: it provides a way of citing the data collection, and it enables a degree of future-proofing by introducing an indirection layer between the identifier and the collection. The re-organisation or movement of collections at a later stage can then take place as long as the owners of the collections undertake to update the relevant persistent identifiers. In the context of supporting and enabling the re-use of data, the persistence provided through these identifiers is crucial to maximizing the length of time during which the data can potentially be re-used. It provides some state to the concept of a data commons. ANDS is working with three different types of identifiers.

The first type is community-assigned identifiers or standards. If a particular discipline community has well-established persistent identifier practices, then ANDS has no interest in replacing these.

The second type is those identifiers provided by ANDS itself. The ANDS Identify My Data service[26] provides persistent identifier minting and management based on the Handles[27] infrastructure. Both human and machine interfaces are available. There are a number of ANDS guides dealing with persistent identifiers at our website[28].

The third type of persistent identifier is the Digital Object Identifier (DOI)[29]. This system, based on the Handles infrastructure, is being increasingly used by publishers to identify publications. ANDS cares about the links between publications and the data collections that underpin them, and also is interested in providing metrics on data citation. ANDS is therefore investigating joining a consortium to enable it to offer DOIs where requested or appropriate.

Register (Supports Objectives E, F)

Within this model, the verb Register pertains to registering collection descriptions and related information (see Describe step above) with one or more public registers of collections. This act of contributing to a larger pool, or making the existence of the data known to a new jurisdiction is the verb Register, and this is an important element of the re-use of data.

There is a spectrum of other ways to approach this Register goal. One end of this spectrum would not even involve a formal registry at all, but would rather involve creating links into the 'semantic web' mesh of linked relationships.[30] These approaches are also counted under Register because they still involve 'registering' the existence of the dataset into a larger pool.

---

[26] Identify My Data overview http://ands.org.au/services/identify-my-data.html
[27] The CNRI Handles system http://www.handle.net/
[28] ANDS Guides http://ands.org.au/guides/
[29] Digital Object Identifier homepage http://www.doi.org/
[30] See the Linked Data Initiative http://linkeddata.org/

ANDS does not presently offer any services to support the semantic web end of the Register spectrum.[31] ANDS does however run a formal register of research data collections. A collections registry is an application set up to harvest these authoritative descriptions and make them available to a variety of browse, discovery, query, and search environments. A collections/service registry is a brokerage service that has enough information about access services and protocols to facilitate automated access to collections and enable machine to machine workflows.

Within the context of ANDS operations, once a data collection has been described, the description needs to be registered in the ANDS Collections Registry. There are three main ways in which this might occur.

The first is for the data owner (or intermediary) to directly add a record using a web interface. This is likely to be of most use where the number of data collections is small and static in nature.

The second way is to store the RIF-CS (or XML to be transformed) on a server at the data provider. ANDS will then issue an HTTP GET to retrieve the XML and load it into the Collections Registry, replacing the previous entries. With this approach, the entire XML package has to be harvested each time, so this is not suitable for lots of collection descriptions.

The third, and preferred, way is to make the collection descriptions harvestable via OAI-PMH[32]. This allows for regular re-harvesting, as well as incremental harvesting (all new records since last harvest). This requires the data source owners to implement an OAI-PMH Data Provider, but there are a number of open source implementations to draw on. It also requires the collection descriptions to be managed as data objects themselves by the data repository.

Discover (Supports Objectives E, F)

The data discovery process has already been outlined above. However, it is worth pointing out that ANDS is providing a range of discovery services, of which the one available at Research Data Australia is just the first phase. These discovery services are provided as part of a discovery ecosystem, shown in Figure 1.

The next phases of the ANDS Research Data Pages will progressively add network graphs, a map interface, tag cloud functionality and advanced search.

In addition, ANDS seeks to interoperate with discipline-specific search services by providing an awareness service for those outside that discipline. In other words, someone might discover through ANDS a dataset in a discipline that they might not have considered. ANDS might then provide them with a link to a more specific search interface provided by someone else that is optimized for that particular data collection.

---

[31] There is no reason however why the ANDS Collections/ Service Registry could not have a SPARQL endpoint, thereby providing an interface into the "linked data" world.
[32] Open Archives Initaitive – Protocol for Metadata Harvesting http://www.openarchives.org/pmh/

Access (Supports Objectives E, F, G)

Once a user has identified a data collection of interest, ANDS will provide information about how to access it.
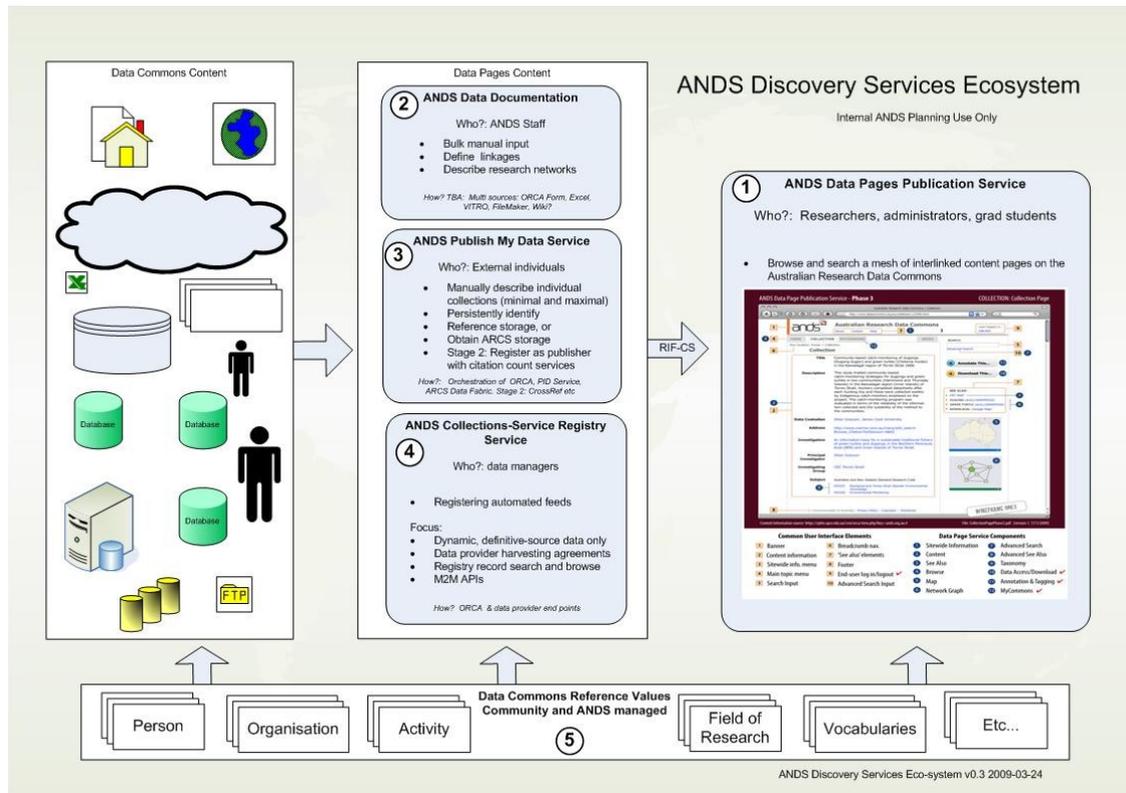

Figure 1. ANDS Discovery Services Ecosystem. Licensed under Creative Commons BY-NC.

In most cases, this will be a link to the underlying data store, allowing a click through. ANDS allows for the description of both open- and closed-access data collections. In the latter case, the access control is enforced by the data store. ANDS does not require a login before searching, and neither does Google (who we anticipate will be the main search mechanism). This means that we are unable to restrict returned search results to only those that the user can access. As we anticipate that the majority of data will be open-access, users should only rarely hit authentication blocks.

In some cases, it will not be possible to gain electronic access directly. This is either because the data are not available in an accessible store (behind a firewall, or not digitized) or because the data owner has requested that any potential users access the data through them. In this case, contact details will be provided in the form of an email address, phone number or postal address. We anticipate that these cases will be in a minority. ANDS currently does not require any specific access policies, although there is a strong encouragement towards open access for publicly funded data.

Exploit (Supports Objective F)

The final Data Verb is Exploit. This is where actual re-use is made of a data object. Services and infrastructure to support Exploit are mostly a very discipline-specific or data-product-specific. Generic national enabling services (of the type ANDS might provide or resource) are less common. The Exploit step is enabled through the existence of good technical metadata (calibrations, classifications, metrics, etc) as well as information about the context of the observation or investigation.

Over the next year ANDS will commission progressively a range of services to exploit the available data fully through data integration/fusion/merging, data visualisation, and data analysis.

The approach taken will be to work closely with leading research groups and Australian Government funded Super-Science[33] initiatives on specific tools for the use and re-use of the newly created pool of data assets. The resulting infrastructure components will depend on the demands of the specific disciplines being targeted, and the data types involved. The overall approach will be based on the exploitation of metadata that are being captured. Rather than designing an architectural approach for particular sets of solutions, ANDS will deliberately first tackle solutions to specific sets of research problems and then generalise the solution.

Next Steps

Better Support For Re-Use

As ANDS has continued to consider what will most enhance the likelihood of re-use, we have drawn on the lessons of the UK Data Archive[34]. Adapting their approach to Data Documentation and Metadata[35], we are now advocating that for the best chance of data re-use we will need to provide:

- Discovery information to enable a user to discover the existence of a data collection. For ANDS this is our RIF-CS description, which enables the user to easily find the data collection through the Research Data Pages that we generate and make available to web search engines. This is what the UKDA calls its catalogue record.
- Assessment information to enable a user to decide if the discovered data are of interest to them. For ANDS this includes the rich context provided by the Research Data Australia pages, but where possible we are also seeking to link to information about the research project that generated the data. This might be something as simple as the original research proposal or the paper that describes the results of the data analysis. This is what the UKDA calls its study-level documentation. The difference is that ANDS will draw on existing information whereas the UKDA has specialist staff who create this.

---

[33] Super Science press release
http://minister.innovation.gov.au/Carr/Pages/SUPERSCIENCEINITIATIVE.aspx
[34] UK Data Archive website http://www.data-archive.ac.uk/
[35] UKDA Data Documentation and Metadata overview
http://www.data-archive.ac.uk/sharing/metadata.asp

- Access information to enable the user to determine how to access the data, and what constraints (if any) might apply to its re-use. ANDS does not have a standard set of access policies, as the underlying data stores are highly heterogenous.
- Re-use information to enable a user to make use of the data once they have decided is of interest to them. This includes, for example, calibration settings, technical metadata, variable names, spreadsheet column explanations and so on. This is what the UKDA calls its data-level documentation. This will usually need to be provided by the researcher or, even better, captured automatically as close as possible to the time of creating (see the Create data verb above).

ANDS is currently determining the best way to represent this rich set of information about a collection. We may expand RIF-CS to accommodate it, or we may decide to implement this augmented collection context using OAI-ORE[36].

Adding a "Define" Verb

As we engage with clients, we are finding that many of the institution repository managers are more familiar with objects and are seeking guidance on collections. This is not just assistance on what ANDS regards as an appropriate collection description, but also assistance with the analysis task of determining what is an appropriate set of collections in their setting. For instance, is the set of ethno-musicology recordings referred to above (30 years of work from one researcher!) just one collection, or a number divided by time, geography or type of performance. In many cases there is no one correct answer, but it is possible that the Define step will need to become a new data verb. This might be seen as a preliminary or pre-requisite part of the Describe activity. Having classification schemes for the values used in descriptions is probably another prerequisite of Describe. ANDS will be partnering with a number of government agencies in Australia to establish and promote infrastructure to support standard classification of key values of interest (place, person, organization, filed of research etc).

Observations

What the Data Verbs Are Not

This paper has been focused heavily on the data verbs that underpin discovery for re-use. This should not be taken to encompass the entirety of ANDS activities. ANDS is also working on data management, funding policies, data capture from a range of sources and much more besides.

The data verbs should also not be taken as yet another attempt at a generic data lifecycle to compete with the DCC Data Curation Lifecycle[37]. The focus of the ANDS data verbs is deliberately on the activities that ANDS needs to undertake, in its national setting, to achieve the data discovery and re-use that is required to meet the vision of more researchers re-using more data more often.

---

[36] Open Archives Initiative – Object Re-Use and Exchange http://www.openarchives.org/ore/
[37] Data Curation Lifecycle Model http://dcc.ac.uk/lifecycle-model/

We are still developing policies for the application of the various verbs (or the specific services that we are building in support of the verbs). In many cases, these are policies that need to be followed by human actors. As a result, we do not yet have any implementation of policies as computer-actionable verbs. The aspects of ANDS described here are not a data management system like IRODS, but are more like a far less ambitious subset of the JISC Information Environment[38]. As the verbs are at a much higher level than something like the IRODS[39] rules, it is unlikely we will ever have a software implementation of these policies.

What the Data Verbs Are

The ANDS Data Sharing Verbs are a structuring device and a high-level architectural approach. Over the course of this year, they have been progressively validated as a useful and powerful way of describing the steps that need to take place in order to realize our vision (and by implication, the systems that need to be in place to support these steps). ANDS staff now use the verbs routinely to describe the activities they are undertaking, and we have adopted them as a structuring technique for our operational planning, and even elicitation of user needs.

ANDS Has a *Broader Program Than Reflected Here*

This paper has been focused heavily on discovery for re-use. This should not be taken to encompass the entirety of ANDS activities. ANDS is also working on data management, funding policies, data capture from a range of sources and much more besides in support of all of the ANDS long-term objectives. Those interested are encouraged to visit the ANDS website for more information.

Conclusion

This document has presented one way of thinking about the things that need to take place to encourage and enable data re-use. Hopefully it will, at a minimum, prompt consideration of the approach ANDS is taking, and perhaps serve as a useful structuring device for others' activities.

The verbs allow a useful focus on the functionality that researchers and research organizations need at their disposal to share and re-use data. This is useful from a planning and design perspective. The verbs encourage a focus on the result rather than the systems. They have proved useful in designing flexible service offerings that allow the orchestration or coordination of services in various ways. The assumption of flexibility and the expectation of heterogeneity are helpful starting points for the design and support of a multi-party, multi-layered activity such as the re-use and sharing of data.

The lessons learned so far from the use of the ANDS Data Sharing Verbs are that they have multiple applications:
- for ANDS staff as a structuring device for much of what ANDS does and how it does it;
- for potential consumers of ANDS services to explain why those services matter and to help them understand how they inter-relate;
- for researchers to explain what needs to happen for them to share their data and re-use others'.

---

[38] JISC IE http://www.jisc.ac.uk/whatwedo/themes/informationenvironment.aspx
[39] Integrated Rule Oriented Data System http://www.irods.org/

While obviously not a detailed technical architecture, they seem to have proven their usefulness over the course of 2009 as ANDS has grown and changed very rapidly. It is hoped that their utility will continue over the next eighteen months as ANDS continues to expand its program of work and add to the richness of its services.

References

Androulakis, S., et al. (2008). Federated repositories of X-ray diffraction images. In *Biological CrystallographyVolume 64, Part 7*. Retrieved August 31, 2009, from
Acta Crystallographica Section D: Biological Crystallography, 2008, Volume D64, pages 810-814.

Australian National Data Service Technical Working Group. *Towards the Australian research data commons*. Retrieved July 24, 2008, from
http://www.pfc.org.au/bin/view/Main/Data

Jantz, R. & Giarlo, M.J. (2005, June). Digital preservation: Architecture and technology for trusted digital repositories. *DLib Magazine, 11 (6)*. doi:10.1045/june2005-jantz. Retrieved August 7, 2009, from
http://www.dlib.org/dlib/june05/jantz/06jantz.html

Rosenthal, D. (2007). *Can we afford to preserve large databases?* Project Briefing: Spring 2007 Task Force Meeting. Retrieved August 7, 2009, from
http://www.cni.org/tfms/2007a.spring/abstracts/PB-can-rosenthal.html

Treloar, A. & Harboe-Ree, C. (2008). Data management and the curation continuum: How the Monash experience is informing repository relationships. *Proceedings of VALA 2008*, Melbourne, February 2008. Retrieved November 10, 2009, from http://www.valaconf.org.au/vala2008/papers2008/111_Treloar_Final.pdf

Treloar, A. & Wilkinson, R. (2008). Rethinking metadata creation and management in a data-driven research world. *Proceedings of IEEE e-Science 2008*, December 2008, Indianapolis. Retrieved November 10, 2009, from
http://doi.ieeecomputersociety.org/10.1109/eScience.2008.41