

Publish My Data: Brought to you by the services of ANDS and ARCS, and the letter D.

Dr Adrian Burton¹, Dr Andrew Treloar²

¹Australian National Data Service, Canberra, Australia, adrian.burton@ands.org.au

²Australian National Data Service, Melbourne, Australia, andrew.treloar@ands.org.au

INTRODUCTION

The Australian National Data Service (ANDS) formally commenced 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, 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. (Towards the Australian Data Commons, p. 6)

To deliver against these objectives, ANDS has four inter-related programs of activity (Developing Frameworks, Providing Utilities, Seeding the Commons, Building Capabilities). ANDS also funds 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).

The ANDS vision for the Australian research-producing sector can be summarized as *more researchers re-using and sharing more data more often*.

ANDS DISCOVERY

In order for researchers to re-use and share data, they need to know that it exists. Discovering research data collected and maintained with public funding is one of the core aims of ANDS as articulated in *Towards the Australian Data Commons*. It is a fundamental requirement for ANDS to ensure that researchers be provided with high quality discovery over and access to research data in a form that makes sense to them. To that end, ANDS is committed to building some pieces of the discovery puzzle, while others have already been built, and other pieces again are being built for specific disciplines through other parts of NCRIS, and in some cases through specific NeAT projects.

Sometimes, researchers will gain access directly to the data (most likely when the data is already well understood). In other instances the access may go via information about that data (perhaps described in its metadata, or a relevant document) or to the collection within which the data sits. These latter cases will provide the searcher with a broader perspective on the potential relevance of the data.

In order for ANDS to achieve all this, the data needs first to be stored somewhere. Despite the references to research repositories and data management in objectives A, B and C above, ANDS is not funded to provide data repositories. This was a deliberate decision of the funding agency, and was made an implementation requirement because there already existed at least two possible solutions that provided data storage or could be used to provide it. These were institutional repositories (installed at all Australian universities under the ASHER program in 2007 and 2008) and the data fabric being built out by the Australian Research Collaboration Service (ARCS). During the ANDS Establishment Project, led by one of us (Treloar), the decision was made for ANDS to work closely on the co-ordination of ANDS and ARCS services to ensure that users receive an experience that is both positive and as seamless as possible. This paper describes one of the first fruits of this activity – a combined service called Publish my Data.

PUBLISH MY DATA

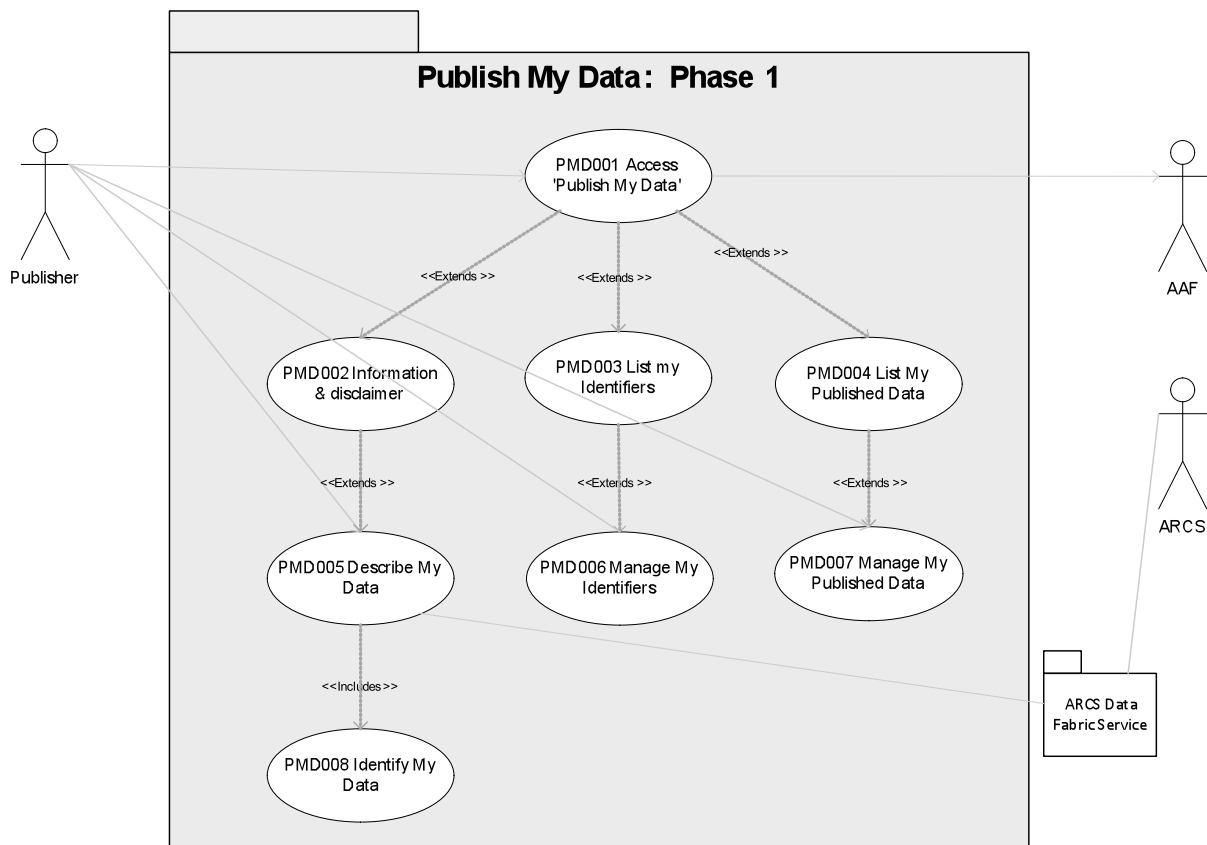
Phase 1 of Publish My Data is an online service which does not enable machine-to-machine transactions. It provides a GUI interface for use by individual publishers, who will usually be researchers. This service will be rolled out in a

number of phases. Phase 1 (the first offering) is intended to enable Australian researchers to store and publicise the existence of research collections via the internet. This will allow AAF authenticated participants to undertake online data entry of collection description information. This information will be stored in the ANDS Collections Registry and be discoverable through Research Data Australia. The Publish My Data workflow will include the allocation of an ANDS Persistent Identifier to each collection registered. The service will include functionality to allow participants to manage and update collection descriptions and persistent identifiers. The intention is that the Publish My Data service can be initiated from either the ANDS website or the ARCS Data Fabric.

The underlying services that will be used to compose Publish My Data are:

- The ARCS Data Fabric (for storage)
- The AAF (for authentication)
- The ANDS Describe My Data service (for describing collections)
- The ANDS Identify My Data service (for persistent identification)
- The ANDS Register My Data service (for registration of collection descriptions)

The following figure shows the services and dependencies.



<The submitted version of the paper will include screenshots and walkthroughs of the user interface>

The process of creating the Publish My Data service has been informed by a careful analysis of user requirements and usability studies of what researchers do. The resulting service will be progressively augmented and refined over time as it gets used.

CONCLUSION

The Publish My Data service provides an easy-to-use, accessible way for researchers to store and publicise their data. It complements existing and foreshadowed institutional solutions, and demonstrates the value of co-ordinating approaches with ARCS.