

Seeding the Australian Data Commons by Blurring the Boundaries

Anna Shadbolt¹ and Simon Porter²

¹University of Melbourne, Melbourne, Victoria, Australia, annams@unimelb.edu.au

²University of Melbourne, Melbourne, Victoria, Australia, sjporter@unimelb.edu.au

The University of Melbourne is one of the richest sources of research data in Australia making it a highly desirable contributor to Australia's research data fabric. This paper will outline how The University of Melbourne partnered with the Australian National Data Service (ANDS)[1] to test a framework for exposing a number of research data collections from a variety of research communities at the university. It will identify how the project evolved as multiple agendas and services blurred the boundaries, enabling the integration that we had only dreamed possible. Core activities in the project included: refining an internal research information and data registration service; building a data archiving service, and supporting local compliance imperatives. Combined they provided a natural flow of opportunities to expose and populate the Australian Data Commons. The presentation will outline the steps taken to integrate the process and to overcome some of the challenges along the way: at the organisational, departmental and researcher levels.

A SYSTEMATIC APPROACH TO ENSURING RESEARCH COMPLIANCE

When research data is published or referred to in a publication, the University of Melbourne has an obligation to make sure that this data is kept for at least 5 years after the date of publication, and in many cases much longer. These obligations are defined in the University's policy on research data and records management (2005) [2], and the Australian Code for the Responsible Conduct of Research (2007) [3]. Currently, Heads of Department are charged with making sure that research data and records are appropriately managed. However, recent surveys and university-wide engagement have indicated that particularly for digital data:

1. Departments do not have the resources to provide long term archival data storage for their researchers;
2. The procedures used for archiving research data are often not coordinated, and,
3. Researchers are often left to find their own solutions.

To meet its obligations, it was proposed that the University establish:

1. A central long term research data archive service, and
2. A central research registry service to record project and research data information all published research data and records at the University of Melbourne.

The process of establishing the second of these services is at the centre of this presentation. In the following section the underlying principles and process for building and populating the registry will be described.

POPULATING THE CENTRAL RESEARCH REGISTRY

Quality descriptions of research data records require not only descriptions of the research data itself, but also the research project and programs that produced it, along with other information to help aid its discovery such as research classifications. Universities collect and report on a vast amount of this information about its research activity. All too often there are already too many requests for new information about research.

In this project we considered it critical that processes used for collecting information about research datasets leverage existing information that the University already collects about research such as grants and publications. Using this information, we tested how readily we were also able to detect the existence of research data sets, along with the probable associated researchers, project description, departments, and research classifications before individual researchers were directly engaged. Such an approach required command of research administrative datasets collected by the Research Office, but also clever use Library technologies to quickly source and scan publications for descriptions of research data.

By leveraging existing information we were able to facilitate the process of research data discovery and populate the central registry. Another advantage of leveraging existing administrative data sets to facilitate the discovery of research data was that it is also possible to make an assessment of the completeness of the exercise.

This internal requirement for rich contextual descriptions of research data in our central registry married very well with the schema for exporting information about research data to the ANDS data commons. The movement of this information and data from the local to the National registry was a natural transition at the technical level. Winning the trust of researchers to leap across the boundary was not quite as straight forward. To illustrate this transition we will outline how this trust was built as we attempted to make the boundaries more permeable with particular reference to the profiling for one of our virtual research institutes.

PROFILING VIRTUAL INSTITUTES

The University of Melbourne has established a number of virtual research institutes as a way of strengthening cross disciplinary research: building connections where none exist; enhancing fledgling connections and providing leadership when responding to research challenges that transcend individual disciplines. These institutes provide an ideal microcosm of the ANDS data commons with similar goals and aspirations for supporting and building cross disciplinary research capability across the nation.

The presentation will close with an outline of our engagement with the Melbourne Sustainable Society Institute [4] highlighting how this activity provided the perfect content and context for testing the agility of the platform to discover research data sets and metadata while building trust in a process that can be of benefit to both individual researchers and Institutes by promoting interdisciplinary collaboration and exposure to the world at large.

REFERENCES

1. More information about the Australian National Data Service can be accessed at: <http://www.ands.org.au/>
2. The University of Melbourne Policy on the Management of Research Data and Records (2005) can be accessed at: <http://www.unimelb.edu.au/records/research.html>. This policy is under review in 2009.
3. The Australian Code for the Responsible Conduct of Research (2007) can be accessed at: <http://www.nhmrc.gov.au/publications/synopses/r39syn.htm>
4. More information about the Melbourne Sustainable Society Institute can be accessed at: <http://www.sustainable.unimelb.edu.au/>