Seeding the Australian Data Commons by blurring the boundaries

Simon Porter and Anna Shadbolt
University of Melbourne

Overview

• Seeing the boundaries
• Seeing the possibilities
• Engaging researchers
• Research data registry
• The registry in action
• Building identity – Internal and external drivers
• Research reporting
• Institutional marketing
• HR data
• Finance data
• Research data compliance
• Research data storage
• Public researcher identity
• Researcher collaboration
• Data commons

• Policy – 2005, review 2009
  – Good start but not enough
• Audit – 2007, 2009
  – Policy and compliance is off the radar for most
  – Compliance – patchy
  – Storage – ad hoc
  – Local registers
  – Risk management
Central resources are the missing key to enabling compliance

- **Central storage** – a petabyte-scale Research Data Centre for Parkville Precinct

- A **Central research data registry**

---

**What is a Central Research Data Registry?**

- **Definition** This is an institution-wide record of all research data and records stored in the university. At a minimum, the Central Registry will include a description of the research data and records, the name(s) of associated researchers and projects, the location of the data (digital and analogue), access restrictions, and relocation and disposal schedules.

- **Intended scale** (biggish)
A central research data registry happens in the context of a richly described research context:

- Grants, Publications, Research Program, External and Internal Researchers, and departments
- All of this information exists and should be reused in the registry...
- Should not be an information silo

---

Piggybacking on our public researcher profile system - ‘find an expert’ dataset...

http://www.findanexpert.unimelb.edu.au/
Bootstrapping the registry

- From Publications to Research data Sets
  - The initial process
  - The output

- Feedback from the Faculty (so far)

Data entry process

1. Pre seed Data Registry with Find an Expert data
2. Scan for references to data in publications
3. Create Skeleton Research Data Record
4. Infer as many of the details about the research data record from the publication
Registry output

Multiple Purposes for the same research data registry

Profiling our research
• Internal drivers
• National drivers
• Global drivers
University Plan 2009: Agree and implement **cross-disciplinary and cross-faculty research themes** for strategic focus and investment as part of a research strategy

**Problem:** How to create identity around a cross disciplinary theme?

---

**Strategy 1**

- Use the cohort of researchers defined by a research theme to define a customized ‘find an expert’ portal…
Strategy 2

- Use the collection of research data ‘assets’ associated with researchers in the theme to aid collaboration discovery and help reinforce identity

External Drivers: ANDS

- In the process of recording research records, we are encouraging researchers to let high level metadata associated with their project be made available to the data commons.
- The case to do this is particularly strong for research themes
- Variable reception across/within disciplines
A coordinated approach to seeding the commons

Challenges

- Current versus eternal research records
- Retrospective task is massive
- Cultural barriers
- Suspicion of role and purpose of a Data Commons
  - Untested
  - Competes with discipline based ‘Commons’
• We have deliberately completed this work with the idea that what we do can be adopted up by other research organisations
• We would love to hear from you if you want to explore options further
  – simon.porter@unimelb.edu.au
  – annams@unimelb.edu.au