The Australian National Data Service – update

David Groenewegen
What is ANDS?

- ANDS is supported by the Australian Government
- Began in 2009, currently funded to June 2013
  - Collaborative Research Infrastructure Scheme (CRIS)
- Collaboration between Monash University, CSIRO and the Australian National University
- Staff in 6 cities across the country
- Funding around 250 projects
Our assumptions:

- Research data is a first class output of research
  - So it should be published
- Research outcomes are enhanced by better access to richer data
  - It’s just like getting access to a more powerful microscope

So better data is a good idea!
ANDS enables transformation of:

Data that are:
- Unmanaged
- Disconnected
- Invisible
- Single use

To Structured Collections that are:
- Managed
- Connected
- Findable
- Reusable

so that Australian researchers can easily publish, discover, access and use research data.
ANDS is helping build the Australian Research Data Commons:

A meeting place for researchers and data:

- The set of data collections that are shareable
- The descriptions of the collections
- The relationships between the data, the researchers, the problems, the instruments and the institutions
- The infrastructure that enables populating and exploiting the commons
ANDS role:

- Establish several national services in support of research data
- Help populate the ARDC with data collections that are managed, connected, and discoverable and reusable
- Partner with institutions to establish coherent institutional research data infrastructure
- Improve the ability of the Australian research system to exploit its research data through the lens of the ARDC using tools, policy and human capability
National Services

- Research Data Australia
- Register My Data enables the harvesting of description information about collections and the activities, parties and services
- Publish My Data allows individuals to manually register a collection description and to obtain a persistent identifier for their collection
- Cite My Data provides cost free creation, resolution and management support for persistent identifiers (DataCite DOIs)
- Describe My Data enables communities to publish and consume existing vocabularies
- Party Infrastructure for independent party (person/organisation) identifies
- Online Gazetteer provides a publicly available interface (both web and machine readable) to an authoritative national gazetteer of place names and other useful spatial information
- Activity Infrastructure persistent identifiers for ARC and NH&MRC research grants
Research Data Australia

- researchdata.ands.org.au
  - 42,000 collections of data
  - associated with 6,000+ linked researchers or research staff
  - harvested through 59 separate data sources
  - and covering all 22 top-level ANZSRC FOR codes.

- 2011 usage: **120,457** page views and **68,031** unique page views over the last year.
- Page views doubled from last year.
ANDS as a funder

- Funding primarily provided at institutional level
- Infrastructure programs
  - Metadata Stores
  - Data Capture
  - Public Sector Data
  - National Collections
  - Applications
- Human capacity programs
  - Seeding the Commons
  - Frameworks & Capabilities
Coherent institutional research data infrastructure

- tools to automatically capture rich metadata along with the data for a wide range of instruments
- operational metadata stores for this metadata
- collections descriptions feeds to ANDS, both Research Institutions and Public Sector data holders
- institutions responsible for nationally significant collections will have been supported in making those collections connected, visible and available, often through an RDSI node
- discipline oriented portals are cross connected to Research Data Australia.
Tools, frameworks and capability

- Improved institutional guidance for internal institutional data management
- Improved institutional guidance for responding to national instruments such as the *Australian Code for the Responsible Conduct of Research*
- Institution wide research data management planning frameworks or improved research data management at many research institutions
- Increased institutional capability for research data management with staff trained in this area
- Tools that enable more effective re-use of research data.
Chinese astronomers have used CSIRO’s Parkes Observatory Pulsar Data Archive to look for glitch events in pulsar data sets.

Results of this work, jointly with CSIRO, are soon to be published.

Archive is also serving as a major resource in an international search for gravitational waves.
Repository of Antibiotic Resistance Cassettes

- database at the University of New South Wales
- instrumental in the discovery of more than 40 new genes in bacteria, including about 10 that may confer antibiotic resistance and are thus clinically important.
Life Patterns

- Project of the University of Melbourne’s Youth Research Centre (edfac.unimelb.edu.au/ycr)
- Online Heritage Resource Manager (OHRM)
- New opportunities:
  - funded by a postdoctoral fellowship from the Swedish government, Dr Evelina Lindstadt will be collaborating with the Life Patterns research team, to undertake a new analysis of the data.
  - The establishment of the OHRM and associated data management processes has also had a demonstration effect, resulting in a successful ARC Linkage project with new partners
Observations of marine organisms by non-scientists can be added to traditional scientific sources.

Mexican graduate student emailed about the possibility of doing similar work in Australia.

Researcher had just picked up an unusual observation of a gloomy octopus (*Octopus tetricus*), traditionally found off Sydney and the east coast of Australia but not in Tasmania.
Australian National Corpus

- New research projects include:
  - Irish influences on Australian English
  - Further developing an interlingua ontology for use across spoken data collections

- The establishment of the AusNC has also seeded collaborative projects including:
  - the development of a virtual Human Science Communication lab
  - the deployment of language data from the AusNC in educational contexts
Going forward

- Exploring linkages of research data with IRs
- Building our relationships with existing partners towards Better Data
- Expanding the use of developed services and tools
- Maintaining what we have built
- Advocacy within the sector
- International relations
Thank you

- Questions?

David.Groenewegen@ands.org.au

ANDS is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy Program and the Education Investment Fund (EIF) Super Science Initiative