ANDS at a glance

- In operation since 2009
- Currently funded by Commonwealth Government under the National Collaborative Research Infrastructure Strategy (NCRIS)
- Approximately $90M received in Commonwealth Funding
- 42 staff (Melbourne, Canberra, Sydney, Brisbane, Adelaide, Perth)
- Successfully completed over $75m (over 200) worth of projects with Universities and PFRO’s across Australia since 2009.
A little bit about ANDS role

The Australian National Data Service (ANDS) is helping through its leadership role, to create a cohesive national collection of research resources and a richer data environment that:

- Makes better use of Australia’s research outputs
- Enables Australian researchers to easily publish, discover, access and use data
- Enables new and more efficient research
In other words...

ANDS Purpose:

To make Australia’s research data assets more valuable for its researchers, research institutions and the nation.
Managing research data...
Too scary?
Too complex?

Simpleinsomnia. (2013). https://farm8.staticflickr.com/7327/11125348744_2a75b75427_z_d.jpg CC By 2.0
Sector Structure

Researchers

Research Centres

Research Institutions

eResearch Units

AURIN
QCIF
MASSIVE
AAP
VeRSI
iMOS
NeCTAR
VERN
iVEC
eRSA
QuestNet
VLSCI
AARNet
Pawsey
SAPAC
TERN
TPAC
NBN
iVEC
NCR
Intersect
EBI Mirror
NCI
BPA
TPAC
ANDS
QRNO
VeRSI
NRN
ANIF
TPAC

Slide taken from the Aero - National Forum of eResearch Service Providers
Defining some key RDM related terms in plain English

“Are you insinuating that I am a purveyor of terminological inexactitudes?”

Winston Churchill
Defining “research data”

“Providing an authoritative definition of research data is challenging, as any definition is likely to depend on the context in which the question is asked.” (ANDS 2014)

More generally, “research data are collected, observed or created, for the purposes of analysis to produce and validate original research results” (DCC)
Research data vary by how they are:

Conceptualised...
- Life sciences
- Physical sciences
- Social sciences
- Humanities
- Arts

Produced...
- Observation
- Experimentation
- Simulation
- Derivation
- Compilation

Stored...
- ASCII
- PDF
- SPSS
- Excel
- PNG
- JPEG
- Java
- XML
- TIFF
- WAVE
- AVI

Represented...
- Text
- Numerical
- Multimedia
- Models
- Software
- Discipline-specific
- Instrument specific
Types of research data
In other words...

Research data are all manner of things produced in the course of research
Defining “data collection” and “dataset”

Generally, not well-defined in the literature, and in some cases there is contention surrounding definitions

- In an RDA context, the terms are somewhat interchangeable, e.g. Collection type might = “collection” or “dataset”
- Terms refer to the type of grouping in which datasets or collections result from
- “Collection” is used as an umbrella term for an aggregation of related datasets or sub-collections
Some common groupings:

- Collections of mixed objects based on a research project

  PhD History project - Interview transcripts and summaries, field notes, personal observations, photographs and digital images

  ECR Toxicology and pharmaceuticals study - Structured data in spreadsheets, databases, experimental observations recorded in lab notebooks
Some common groupings:

- Collections of particular object types based on intellectual themes together with curatorial requirements.
Some common groupings:

- **Collections of digital data**

  Might include scientific observations in a digital format, together with information about scientific equipment and methods used to compile the data.
Some common groupings:

- Collections of digital data or physical objects based on a temporal range such as time series data.
Some common groupings:

- Collections of descriptions (metadata) of one or more collections, parties, activities and services
  
  RDA is an example
In other words...

- A mixed bag of data types based around a project or intellectual theme, are called a “collection”.
- More homogenous data (as in format or type) where the focus is the data, we’d call these “datasets”
- “Collection” is a good term for multiples of related datasets or sub-collections
Defining “data lifecycle”

http://www.data-archive.ac.uk/create-manage/life-cycle
Digital Curation Centre (DCC) – Data lifecycle
ANDS data curation continuum

Private Research Domain
- Laboratory Information Management System/Research Management System
- May link to data objects stored in Research Data Store

Private Research Domain involves the core research team as they undertake the research, usually within a single institution. Access is often tightly controlled as hypotheses and analyses are developed.

Migration Process
- Authorised by research team leader
- Performed by research team IT support

Shared Research Domain
- Collaboration Support System (Plone, TWiki, Sharepoint)
- May link to data objects stored in Collaboration Data Store

Shared Research Domain involves researchers outside the core team as they collaborate with colleagues, often across institutions. Access is more open, but not everything is shared.

Publication Curation Boundary
- May link to data objects stored in Public Data Store/Repository

Public Domain
- Institutional document repository

Public Domain involves the public sphere (publication in the sense of making public). Access will usually be open to all.

Research lifecycle - JISC
In other words...

The data lifecycle identifies the stages that data will pass through and describes the transformations that occur at each stage.
Australian National Data Service

ANDS is enabling the 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
Defining “research data management”

“... the active management and appraisal of data over the lifecycle of scholarly and scientific interest” (DCC)

"Research data management concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information."

RDM involves some high-level questions

- How does the researcher plan to manage their research data?
- What data will be created/colllected/compiled? And how?
- What documentation and metadata will accompany the data?
- How will ethical and/or intellectual property rights issues be managed?
- How will the data be stored and backed up?
- How will access to and security of the data be managed?
- Which data are of long-term value (for sharing and preservation)?
- How will data be shared?
- What is the long-term preservation plan for the data (dataset)?
In other words...

RDM = Taking due care of research data from creation through to long-term preservation or secure disposal
Defining “data sharing”

“Data sharing is the release of research data for use by others. Release may take many forms, from private exchange upon request to deposit in a public data collection. Posting datasets on a public website or providing them to a journal as supplementary materials also qualifies as sharing.”

Sharing research data with collaborators during the project

- Networked drives
- Secure data transfer
- Access controls, where required
- Collaboration spaces and tools
Sharing research data and metadata with wider audiences post project

- Use of appropriate repositories, data journals, websites
- Explicit statements on access conditions: open, conditional, restricted
- Considerations on restrictions to sharing: confidentiality, consent agreements, Copyright and other IP issues
- Explicit conditions for reuse – licensing data
- Clear indications on how to cite the data
In other words...

Sharing research data means using effective mechanisms for dissemination...
Defining “open data”

“Open data are the building blocks of open knowledge. Open knowledge is what open data becomes when it’s useful, usable and used.

The key features of openness are:

- Availability and access
- Reuse and redistribution
- Universal participation”
ANAND projects
✓ Major Open Data Collections (MODCs)
✓ Open Data Collections (ODCs)
In other words...

Value is evident in data that:

- Can be used later
- Are able to be used by more researchers
- Are able to be used to answer new questions
- Are able to be integrated to explore new data spaces

...To do so, data must be managed, connected, discovered, and then re-used – data have to move out of the “lab”
Defining Library RDM roles

- Taking a lead on local (institutional) research data policy and governance
- Bringing data into teaching and learning for students
- Teaching “data literacy” to postgraduate students
- Developing researcher data awareness
- Providing advice, e.g. on planning for data management or on RDM within a project
- Explaining the impact of sharing data, and how to cite data
- Developing a referral service - who in the Uni to consult in relation to a particular question
- Auditing to identify data sets for archiving or RDM needs
- Developing and managing access to data collections
- Documenting what datasets an institution has
- Developing local data management capacity
- Promoting data reuse by highlighting what is available
My aim…

“Simplicity is about subtracting the obvious and adding the meaningful.”

Help from ANDS

- **Guides on the ANDS website**
- **Contact your ANDS Outreach Officer**
- **ANDS run workshops/seminars**
- **ANDS webinars ([YouTube channel](https://www.youtube.com))**
- **Register for [andsUP](https://ands-up.com)***
Thank you!
Acknowledgements

Ideas and content have been taken from various sources:


- DaMaRo Project (2013). Introduction to research data management. http://damaro.oucs.ox.ac.uk/training_materials.xml


- Research Lifecycle at UCF http://library.ucf.edu/ScholarlyCommunication/ResearchLifecycleUCF.php
Acknowledgements

Images

Types of data slide:


- ANDS curation continuum http://ands.org.au/assets/images/curation.continuum.gif
- Bulb-on http://www.salesenlightenment.com/images/bulb_on.jpg
- Lifecycle webDCC http://www.dcc.ac.uk/sites/default/files/lifecycle_web.png
- Produced https://encrypted-tbn0.gstatic.com/images?q=tbn:ANd9GcRESNMvW44FJQ0x7VJ_L3mnRW5eHhljTevpREuK6Byrk4cP0QVYw
- Research Lifecycle ashx http://www.jisc.ac.uk/whatwedo/campaigns/res3/~/media/JISC/campaigns/research/ResearchLifecycle.ashx?w=650&h=752&as=1
- Tango face grin 115990 http://images.all-free-download.com/images/graphiclarge/tango_face_grin_115990.jpg
- UFC Cycle800 http://library.ucf.edu/ScholarlyCommunication/images/Cycle800.jpg
This work is licensed under a Creative Commons Attribution 3.0 Australia License

NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

ANDS is supported by the Australian Government through the National Collaborative Research Infrastructure Strategy (NCRIS).