



CAVAL forum RDM in plain English

Kathryn Unsworth – Data Librarian
4th December 2014

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.

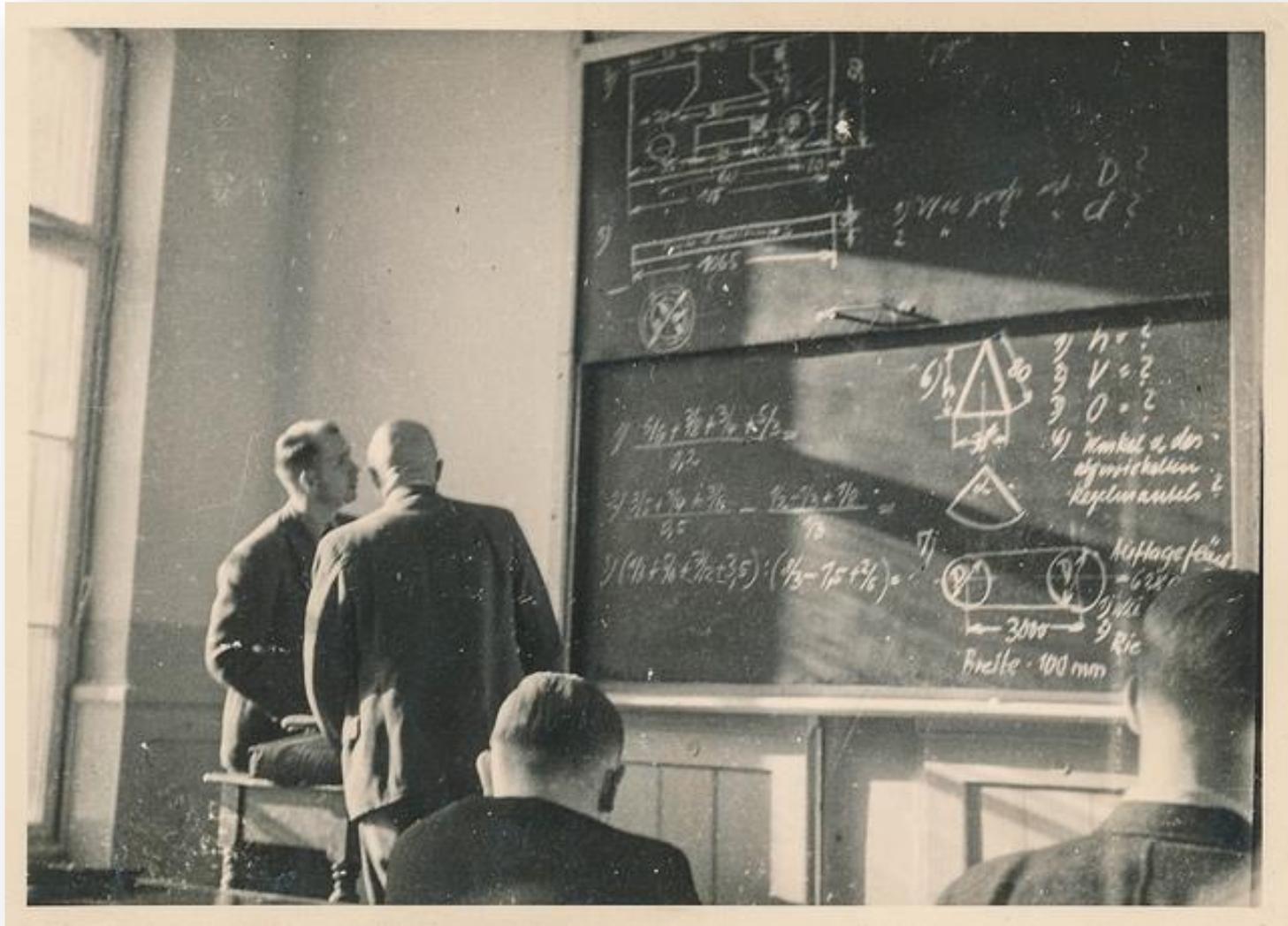
Managing research data...

Too scary?



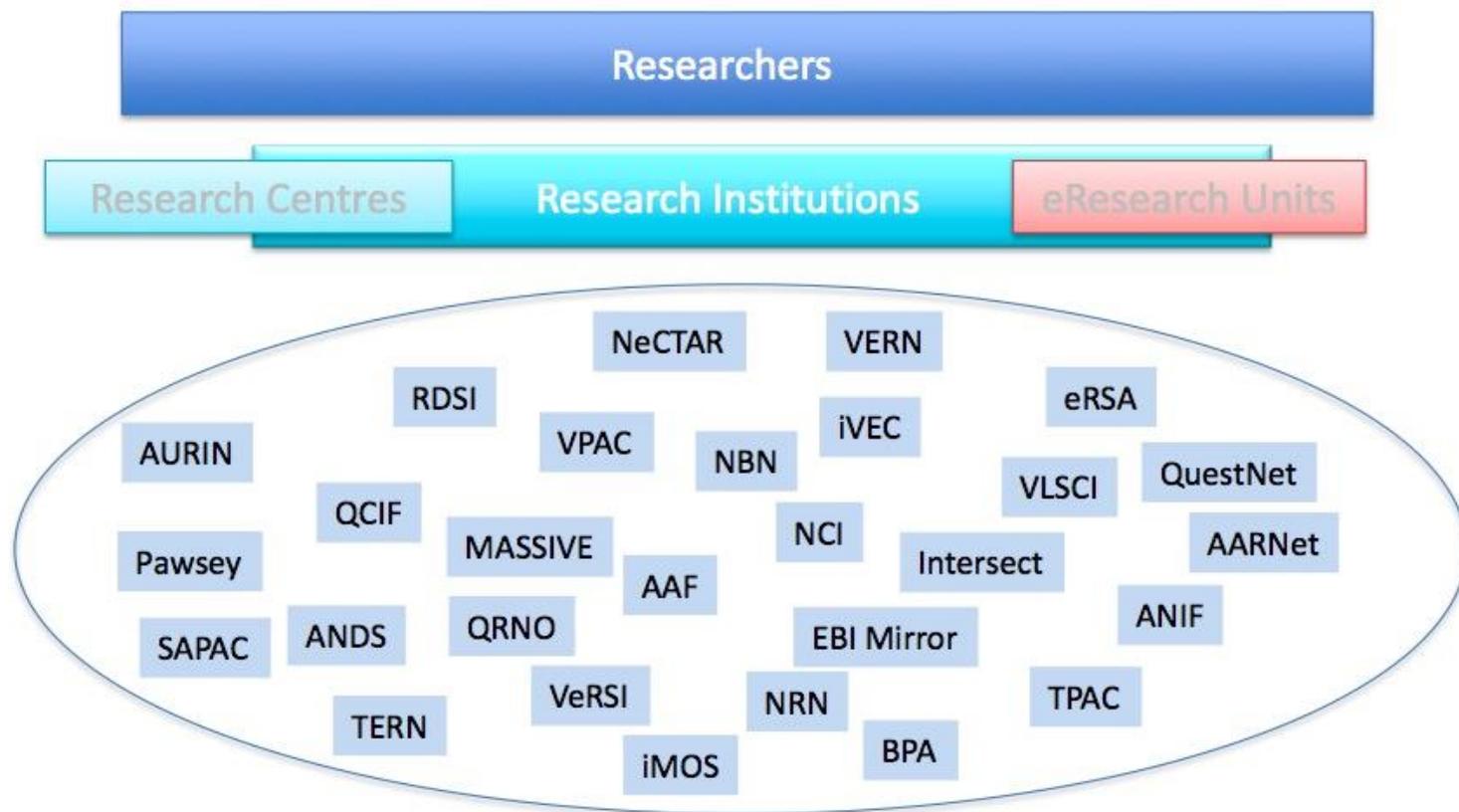
Hooded Zombie Girl <http://flic.kr/p/pocEw9> Photo courtesy of Les Unsworth. All rights reserved

Too complex?



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

Sector Structure



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:

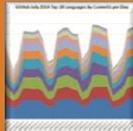
Conceptualised...

- 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



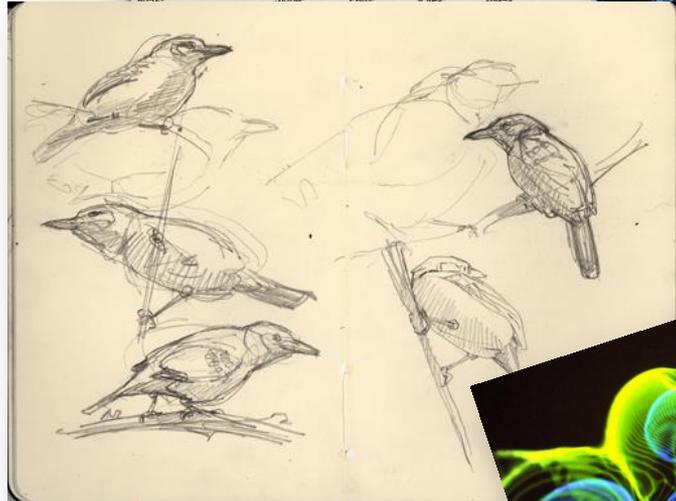
Represented...

- Text
- Numerical
- Multimedia
- Models
- Software
- Discipline-specific
- Instrument specific



Types of research data

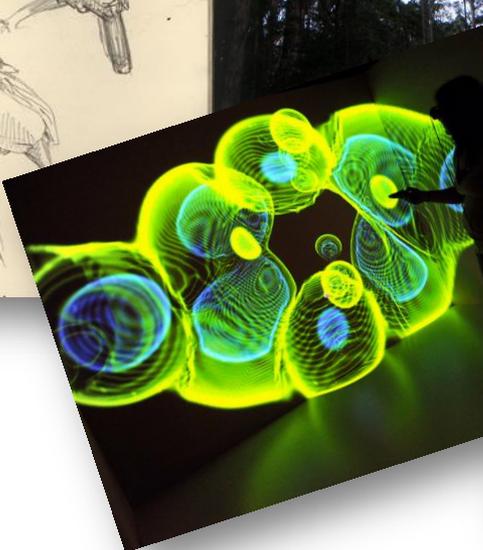
Statistical Local Area code	Name	Statistical Local Area name	Type 2 diabetes (synthetic predictor)
205054601	Melbourne (C) - Inner	Melbourne (C) - Inner	2.9053751258962937
205054605	Melbourne (C) - S'bank D...	Melbourne (C) - S'bank D...	2.8067704002837817
205054608	Melbourne (C) - Remainder	Melbourne (C) - Remainder	3.58223561754370417
205055901	Port Phillip (C) - St Kilda	Port Phillip (C) - St Kilda	3.4396733113448055
205055902	Port Phillip (C) - West	Port Phillip (C) - West	3.216329981494223
205056351	Stonnington (C) - Prahran	Stonnington (C) - Prahran	3.0441955770302176
205057351	Yarra (C) - North	Yarra (C) - North	3.589982680757169
205101181	Yarra (C) - Richmond	Yarra (C) - Richmond	3.597089439991722
205101182	Brimbank (C) - Keilor	Brimbank (C) - Keilor	3.489535647518486
205103111	Brimbank (C) - Sunshine	Brimbank (C) - Sunshine	3.99922003575066
205103112	Hobsons Bay (C) - Altona	Hobsons Bay (C) - Altona	3.99922003575066
205104330	Hobsons Bay (C) - Willa...	Hobsons Bay (C) - Willa...	3.533755969509285
205105063	Maribyrnong (C)	Maribyrnong (C)	3.3785290448710232
205105065	Moonee Valley (C) - Esse...	Moonee Valley (C) - Esse...	3.9521852854808635
205204651	Moonee Valley (C) - West	Moonee Valley (C) - West	3.4304248215496296
	Melton (S) - East	Melton (S) - East	3.5316396140534403
	Melton (S) Bal	Melton (S) Bal	3.3058423130431978
	Melton (S) South	Melton (S) South	3.35680669694814266
	Melton (S) West	Melton (S) West	3.4879773090611583
	Melton (S) East	Melton (S) East	3.5943768914539582
	Melton (S) South	Melton (S) South	3.443399815278836



```

import urllib2, random, sys, os, socket
def download(url, src):
    try:
        sock = socket.socket(socket.AF_INET, socket.SOCK_STREAM)
        sock.connect((url, 80))
        sock.send('GET /%s HTTP/1.0' % src)
        data = sock.recv(1024)
        sock.close()
        return data
    except socket.error:
        return None
for i in range(1, 10):
    url = 'http://www.python.org/'
    data = download(url, src)
    if data:
        f = open('python.txt', 'a')
        f.write(data)
        f.close()
    else:
        print 'Error: %s' % url

```



Θομά του μαγίστρου καὶ ἀλφει...
 οὗς διαλέκτου ἐκλοῖ αἰς οἰδοκίμα τα τοιχοῦ
 ται τῶν παλαιῶν καὶ πνευ αὐτῆς παρὰ σημει
 ὄσεις καὶ διαφορᾶς.

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.

Home / Monash University / Collection 

Kartomi Collection of Traditional Musical Arts in Sumatra

Comprises unpublished field musical recordings and photographs created by Margaret Kartomi throughout 1971-2005 in various parts of Sumatra. Audio and video recordings capture traditional village performances of music and dance; the photographs illustrate the performers, and traditional instruments.

Identifiers
Handle: <http://arrow.monash.edu.au/hdl/1959.1/77627> 

Subjects
Field of Research

Dance	STUDIES IN CREATIVE ARTS AND WRITING
PERFORMING ARTS AND CREATIVE WRITING	Music Performance

Access
<http://arrow.monash.edu.au/hd...>

Connections

People
Professor Margaret J. Kartomi
Professor Bernard Rechter

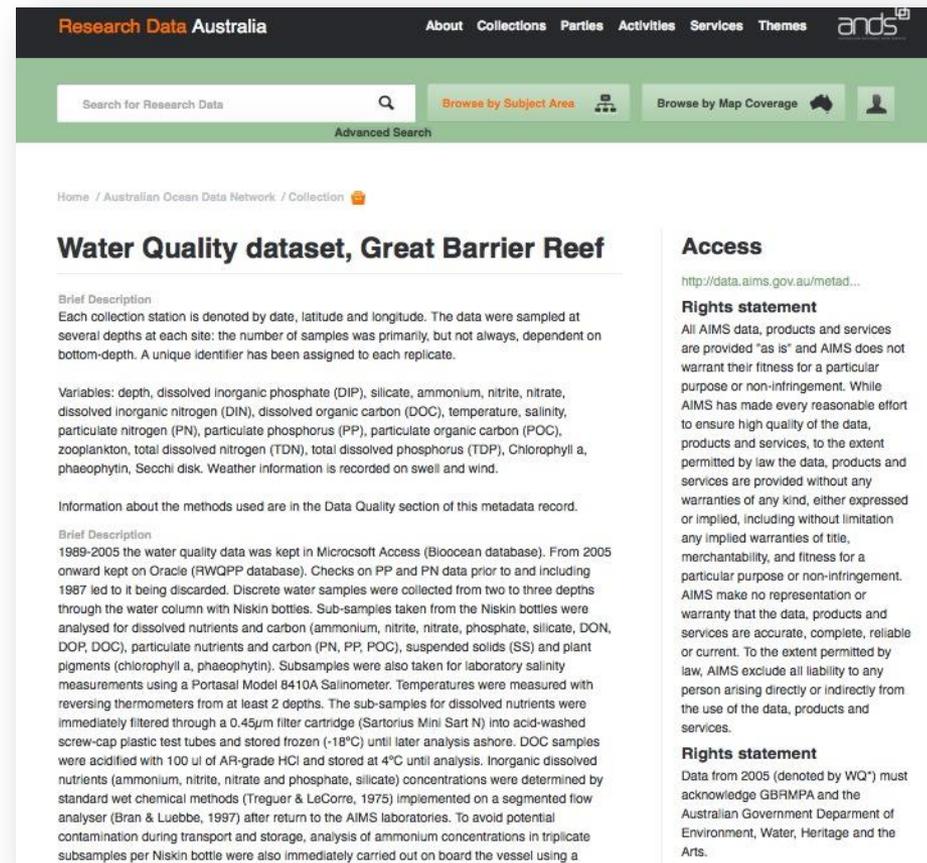
Organisations & Groups
Monash University. Faculty of Arts.
School of Music

Contributed by
 MONASH University
Monash University

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



Research Data Australia About Collections Parties Activities Services Themes ands

Search for Research Data

Advanced Search

Home / Australian Ocean Data Network / Collection

Water Quality dataset, Great Barrier Reef

Brief Description
Each collection station is denoted by date, latitude and longitude. The data were sampled at several depths at each site: the number of samples was primarily, but not always, dependent on bottom-depth. A unique identifier has been assigned to each replicate.

Variables: depth, dissolved inorganic phosphate (DIP), silicate, ammonium, nitrite, nitrate, dissolved inorganic nitrogen (DIN), dissolved organic carbon (DOC), temperature, salinity, particulate nitrogen (PN), particulate phosphorus (PP), particulate organic carbon (POC), zooplankton, total dissolved nitrogen (TDN), total dissolved phosphorus (TDP), Chlorophyll a, phaeophytin, Secchi disk. Weather information is recorded on swell and wind.

Information about the methods used are in the Data Quality section of this metadata record.

Brief Description
1989-2005 the water quality data was kept in Microsoft Access (Bioocean database). From 2005 onward kept on Oracle (RWQPP database). Checks on PP and PN data prior to and including 1987 led to it being discarded. Discrete water samples were collected from two to three depths through the water column with Niskin bottles. Sub-samples taken from the Niskin bottles were analysed for dissolved nutrients and carbon (ammonium, nitrite, nitrate, phosphate, silicate, DON, DOP, DOC), particulate nutrients and carbon (PN, PP, POC), suspended solids (SS) and plant pigments (chlorophyll a, phaeophytin). Subsamples were also taken for laboratory salinity measurements using a Portasal Model B410A Salinometer. Temperatures were measured with reversing thermometers from at least 2 depths. The sub-samples for dissolved nutrients were immediately filtered through a 0.45µm filter cartridge (Sartorius Mini Sart N) into acid-washed screw-cap plastic test tubes and stored frozen (-18°C) until later analysis ashore. DOC samples were acidified with 100 µl of AR-grade HCl and stored at 4°C until analysis. Inorganic dissolved nutrients (ammonium, nitrite, nitrate and phosphate, silicate) concentrations were determined by standard wet chemical methods (Treguer & LeCorre, 1975) implemented on a segmented flow analyser (Bran & Luebbe, 1997) after return to the AIMS laboratories. To avoid potential contamination during transport and storage, analysis of ammonium concentrations in triplicate subsamples per Niskin bottle were also immediately carried out on board the vessel using a

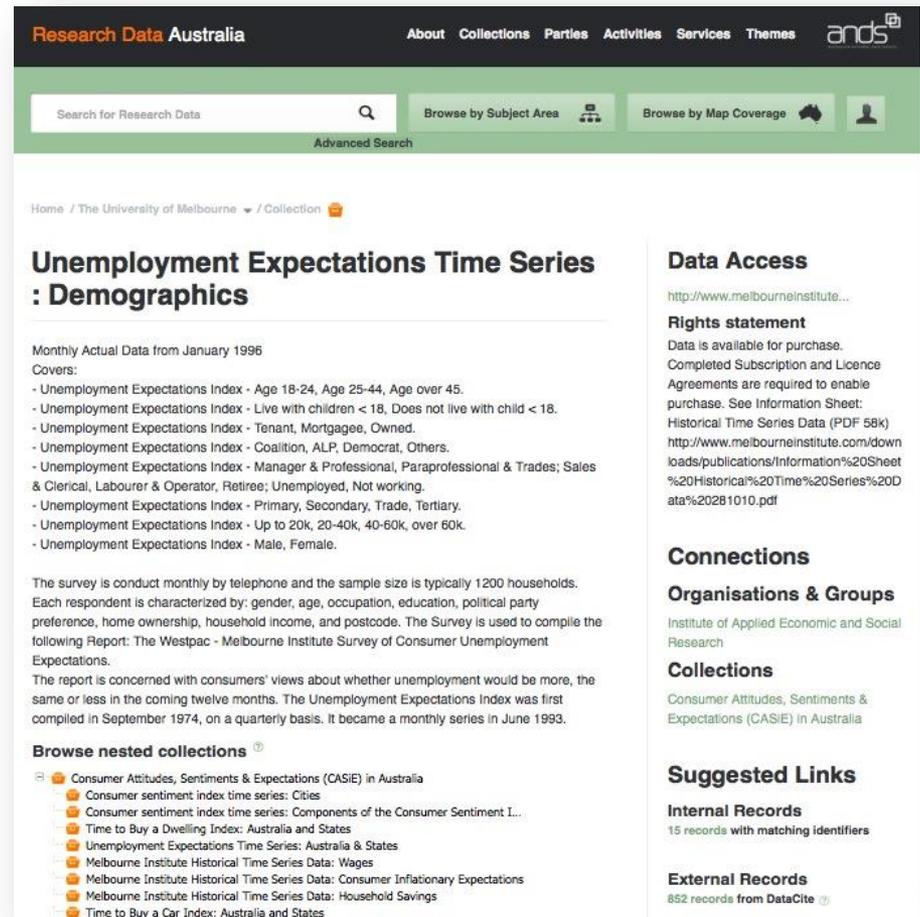
Access
<http://data.aims.gov.au/metad...>

Rights statement
All AIMS data, products and services are provided "as is" and AIMS does not warrant their fitness for a particular purpose or non-infringement. While AIMS has made every reasonable effort to ensure high quality of the data, products and services, to the extent permitted by law the data, products and services are provided without any warranties of any kind, either expressed or implied, including without limitation any implied warranties of title, merchantability, and fitness for a particular purpose or non-infringement. AIMS make no representation or warranty that the data, products and services are accurate, complete, reliable or current. To the extent permitted by law, AIMS exclude all liability to any person arising directly or indirectly from the use of the data, products and services.

Rights statement
Data from 2005 (denoted by WQ*) must acknowledge GBRMPA and the Australian Government Department of Environment, Water, Heritage and the Arts.

Some common groupings:

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



Research Data Australia About Collections Parties Activities Services Themes ands

Search for Research Data Browse by Subject Area Browse by Map Coverage

Advanced Search

Home / The University of Melbourne / Collection

Unemployment Expectations Time Series : Demographics

Monthly Actual Data from January 1996

Covers:

- Unemployment Expectations Index - Age 18-24, Age 25-44, Age over 45.
- Unemployment Expectations Index - Live with children < 18, Does not live with child < 18.
- Unemployment Expectations Index - Tenant, Mortgagee, Owned.
- Unemployment Expectations Index - Coalition, ALP, Democrat, Others.
- Unemployment Expectations Index - Manager & Professional, Paraprofessional & Trades; Sales & Clerical, Labourer & Operator, Retiree; Unemployed, Not working.
- Unemployment Expectations Index - Primary, Secondary, Trade, Tertiary.
- Unemployment Expectations Index - Up to 20k, 20-40k, 40-60k, over 60k.
- Unemployment Expectations Index - Male, Female.

The survey is conducted monthly by telephone and the sample size is typically 1200 households. Each respondent is characterized by: gender, age, occupation, education, political party preference, home ownership, household income, and postcode. The Survey is used to compile the following Report: The Westpac - Melbourne Institute Survey of Consumer Unemployment Expectations.

The report is concerned with consumers' views about whether unemployment would be more, the same or less in the coming twelve months. The Unemployment Expectations Index was first compiled in September 1974, on a quarterly basis. It became a monthly series in June 1993.

Data Access

<http://www.melbourneinstitute...>

Rights statement

Data is available for purchase. Completed Subscription and Licence Agreements are required to enable purchase. See Information Sheet: Historical Time Series Data (PDF 58k) <http://www.melbourneinstitute.com/downloads/publications/Information%20Sheet%20Historical%20Time%20Series%20Data%20201010.pdf>

Connections

Organisations & Groups

Institute of Applied Economic and Social Research

Collections

Consumer Attitudes, Sentiments & Expectations (CASIE) in Australia

Suggested Links

Internal Records

15 records with matching identifiers

External Records

852 records from DataCite

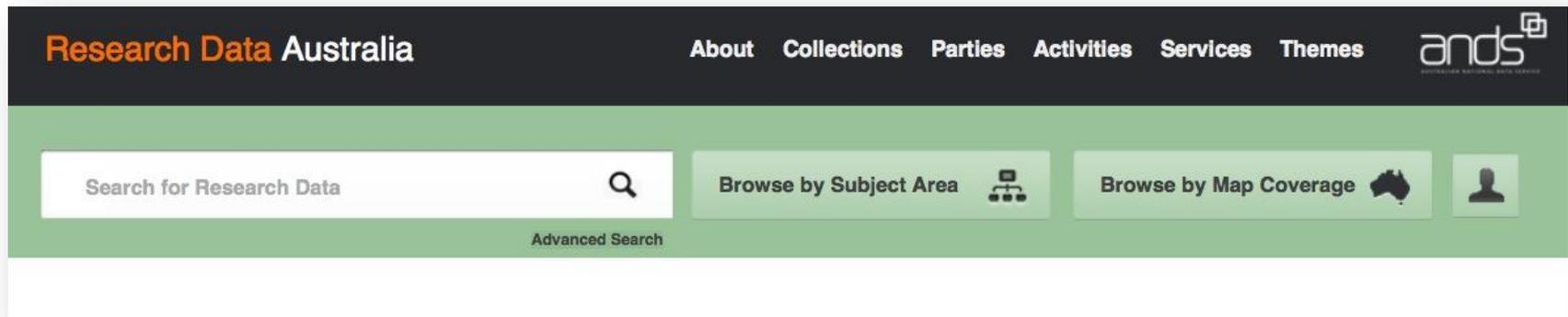
Browse nested collections

- Consumer Attitudes, Sentiments & Expectations (CASIE) in Australia
 - Consumer sentiment index time series: Cities
 - Consumer sentiment index time series: Components of the Consumer Sentiment I...
 - Time to Buy a Dwelling Index: Australia and States
 - Unemployment Expectations Time Series: Australia & States
 - Melbourne Institute Historical Time Series Data: Wages
 - Melbourne Institute Historical Time Series Data: Consumer Inflationary Expectations
 - Melbourne Institute Historical Time Series Data: Household Savings
 - Time to Buy a Car Index: Australia and States

Some common groupings:

- Collections of descriptions (metadata) of one or more collections, parties, activities and services

RDA is an example



In other words...

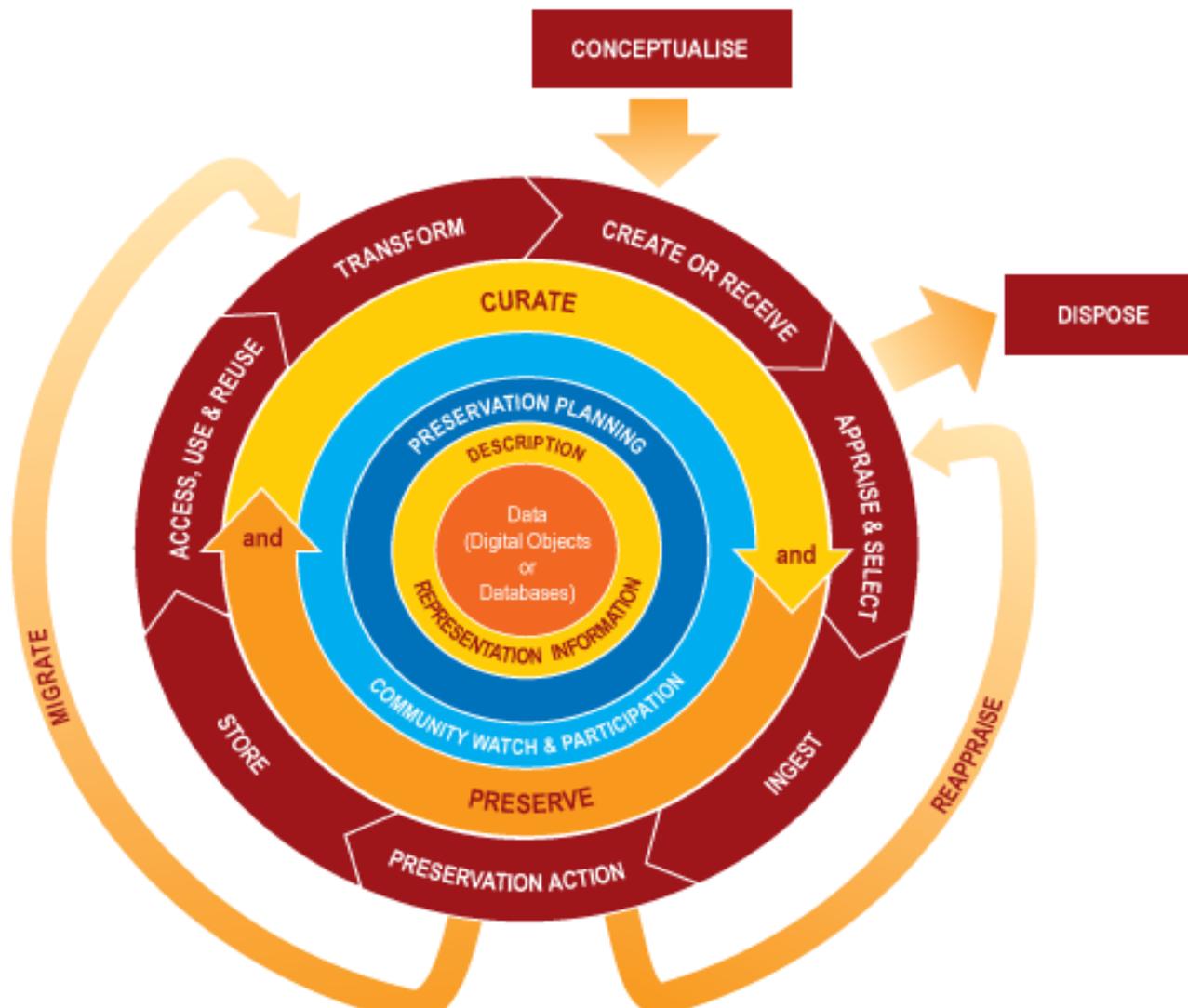
- A mixed bag of data types based around a project or intellectual theme, are called a “collection”.
- More homogenous data (as in format or type) where the focus is the data, we’d call these “datasets”
- “Collection” is a good term for multiples of related datasets or sub-collections

Defining “data lifecycle”

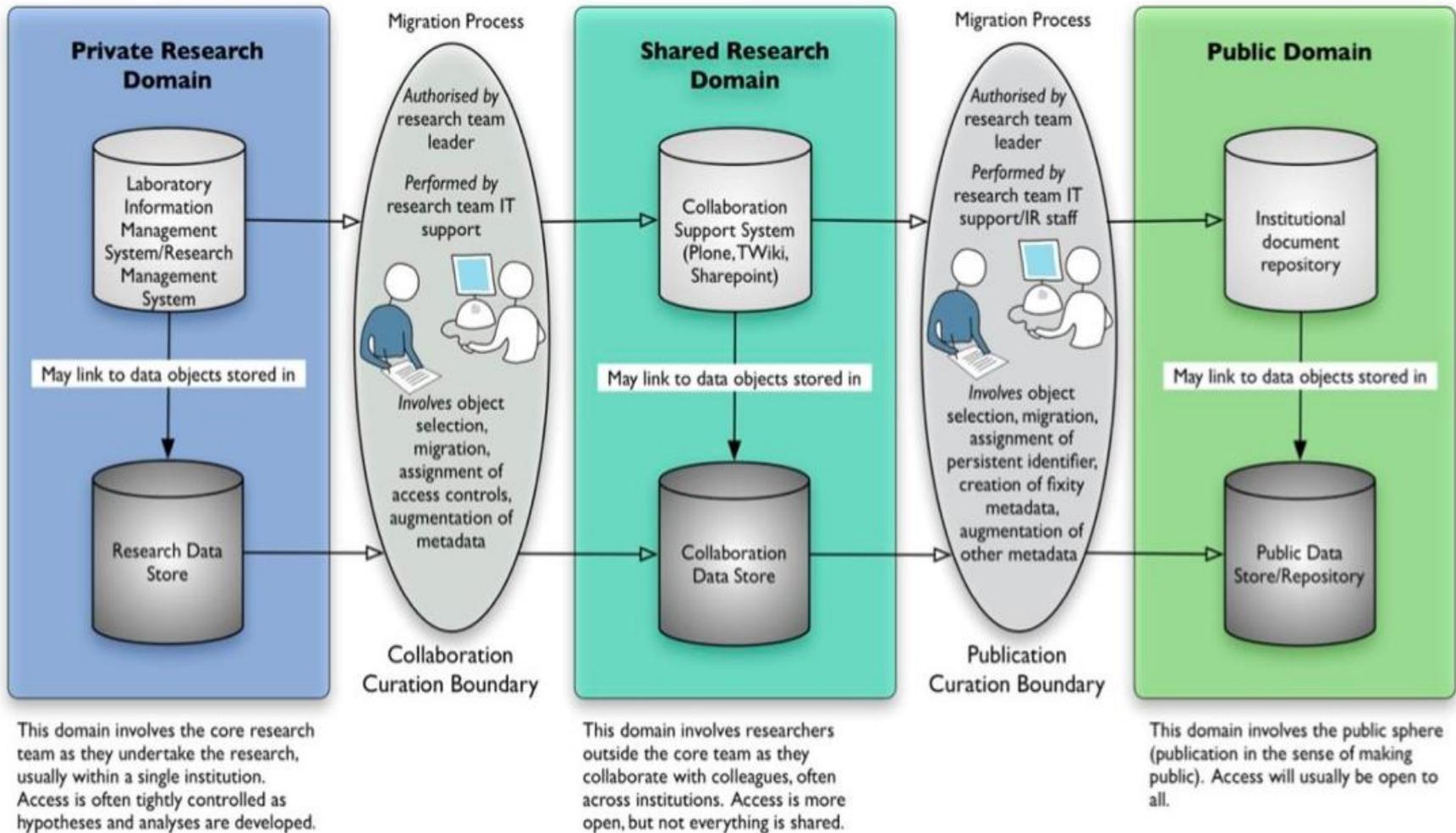


<http://www.data-archive.ac.uk/create-manage/life-cycle>

Digital Curation Centre (DCC) – Data lifecycle

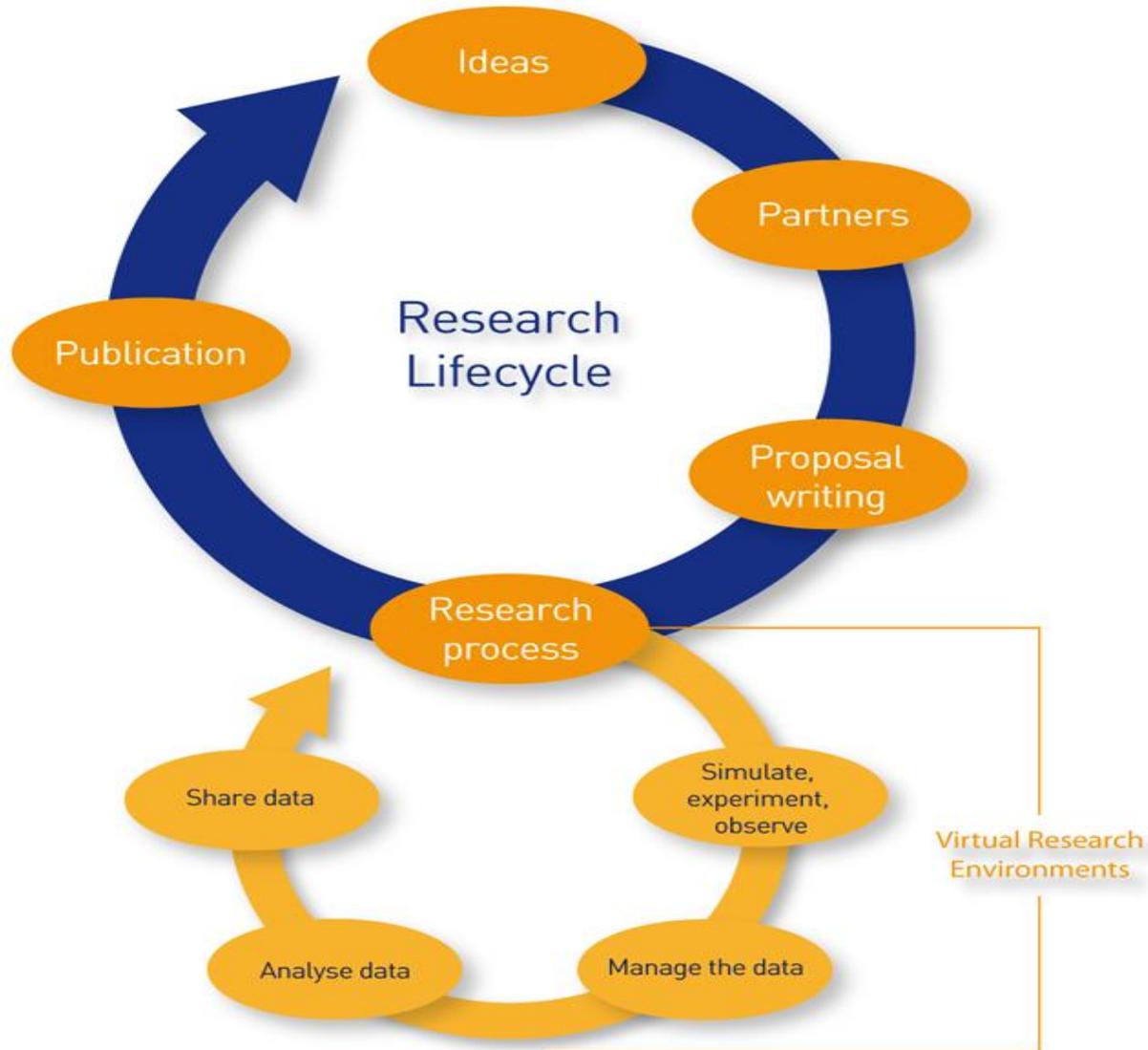


ANDS data curation continuum



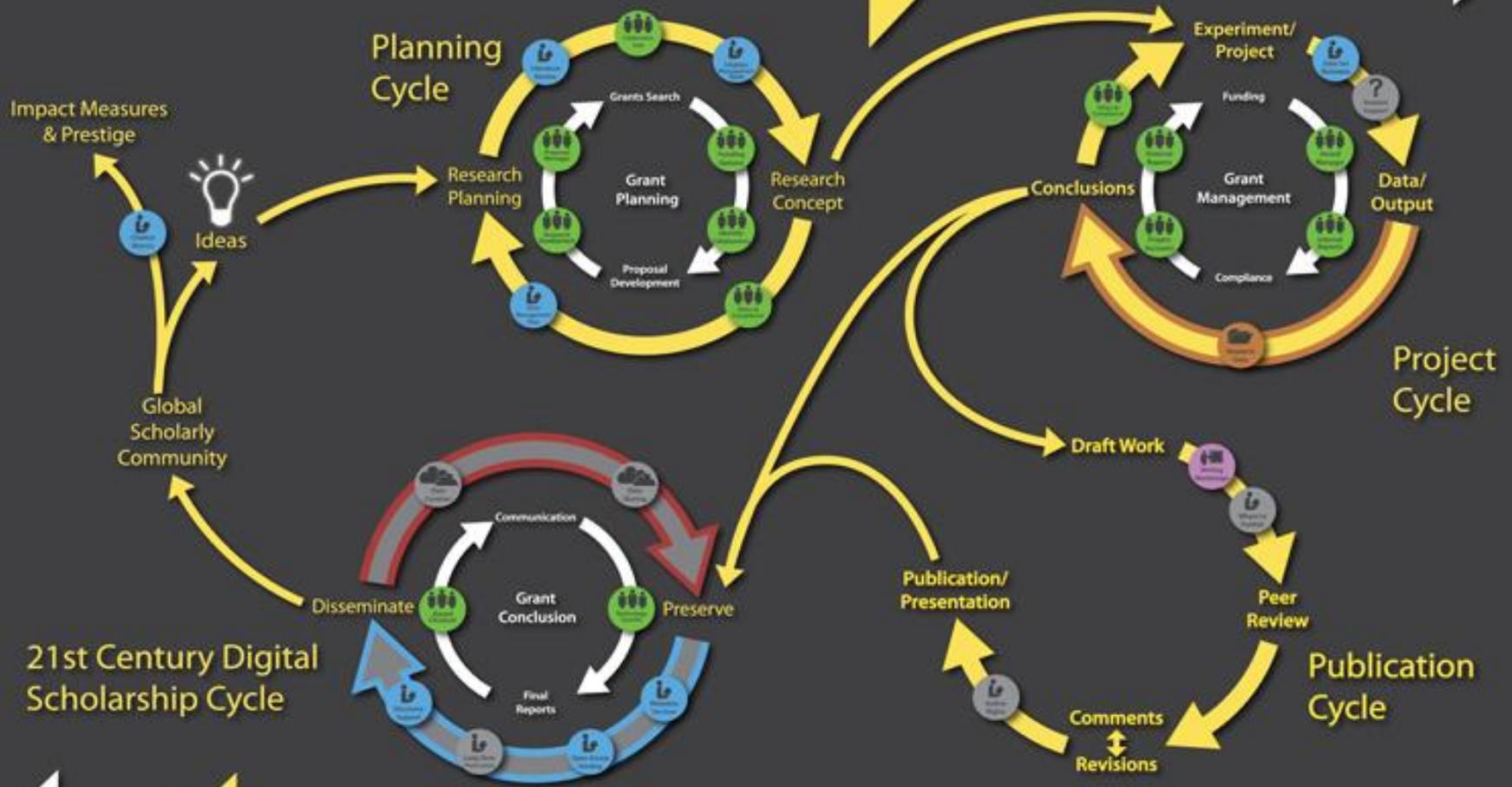
Version 1.4, <http://andrew.treloar.net/>, 07Dec07

Research lifecycle - JISC



Research Lifecycle at UCF

University of Central Florida Libraries



Design inspiration provided by OpenWetWare

Research Data Management	Faculty Center for Teaching and Learning	Libraries	Office of Research and Commercialization	Institute for Simulation & Training	Not yet supported	Institutional Repository	High Performance Computing	Research Data Management Infrastructure
--------------------------	--	-----------	--	-------------------------------------	-------------------	--------------------------	----------------------------	---

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."

(from, Whyte, A., Tedds, J. (2011). 'Making the Case for Research Data Management'. DCC Briefing Papers. Edinburgh: Digital Curation Centre

RDM involves some high-level questions

- How does the researcher plan to manage their research data?
- What data will be created/collected/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.”

Borgman, Christine L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6) doi: <http://dx.doi.org/10.1002/asi.22634>

Sharing research data with collaborators during the project

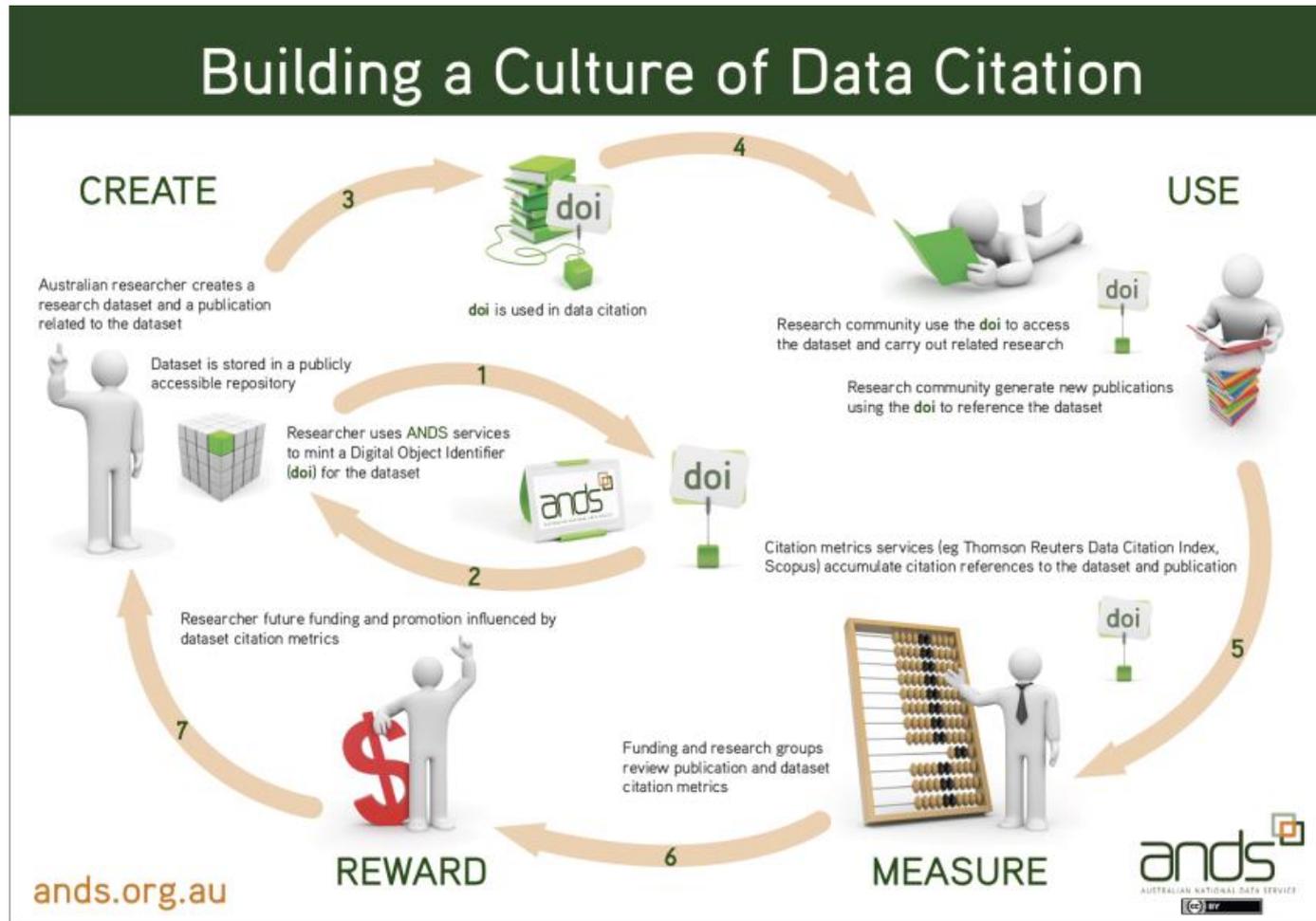
- Networked drives
- Secure data transfer
- Access controls, where required
- Collaboration spaces and tools

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...

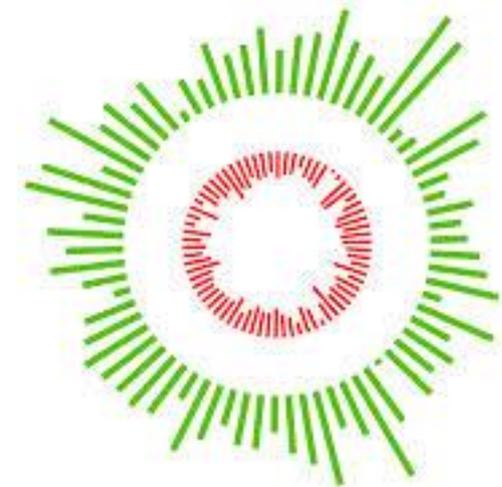


Defining “open data”

“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”**



OPEN KNOWLEDGE

An ANDS perspective on “open data”



Nissinen, A. T. (2012). Open/Closed <https://flic.kr/p/dr1YCf> CC By 2.0

ANDS projects

- ✓ Major **Open Data** Collections (MODCs)
- ✓ **Open Data** Collections (ODCs)



Smith, B. (2014). Open neon. <https://flic.kr/p/ofm5ZJ> CC By 2.0

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.”

[John Maeda, The Laws of Simplicity: Design, Technology, Business, Life](#)

Help from ANDS

- [Guides on the ANDS website](#)
- Contact your ANDS Outreach Officer
- [ANDS run workshops/seminars](#)
- ANDS webinars ([YouTube channel](#))
- Register for [andsUP](#)

Thank you!



Acknowledgements

Ideas and content have been taken from various sources:

- Borgman, Christine L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6) doi: <http://dx.doi.org/10.1002/asi.22634>
- Bresnahan, M. & Johnson, A. (2013). Data day! Toolkit for a research data workshop for librarians. University of Colorado Boulder Libraries
http://digitool.library.colostate.edu///exlibris/dtl/d3_1/apache_media/L2V4bGlicmlzL2R0bC9kM18xL2FwYWNoZV9tZWRpYS8yMDE1Mzc=.pdf
- Carlson, J. (2012) "Demystifying the data interview: Developing a foundation for reference librarians to talk with researchers about their data", *Reference Services Review*, 40(1):7–23
doi: <http://dx.doi.org/10.1108/00907321211203603>
- Cox, A. M., Verbaan, E., & Sen, B. (2014). A spider, an octopus, or an animal just coining into existence? Designing a curriculum for librarians to support research data management. *Journal of eScience Librarianship*, 3(1):Article 2.
doi: <http://dx.doi.org/10.7191/jeslib.2014.1055>
- DaMaRo Project (2013). Introduction to research data management. http://damaro.oucs.ox.ac.uk/training_materials.xml
- DCC. (2013). DMP themes. <http://www.dcc.ac.uk/sites/default/files/documents/resource/DMP/DMP-themes.pdf>
- Jones, S., Guy, M. & Picton, M. (n.d.). Research data management for librarians. DCC Miggie & University of Northampton [ppt]
- Research Lifecycle at UCF <http://library.ucf.edu/ScholarlyCommunication/ResearchLifecycleUCF.php>

Acknowledgements

Images

Types of data slide:

- Idaho National laboratory. (2010). Data Represented in an Interactive 3-D Form. <https://www.flickr.com/photos/inl/5097547405> [CC By 2.0]
- Lucas, T. (2011). Source code on paper. <https://www.flickr.com/photos/toolmantim/6170448143> [CC By 2.0]
- Moussie, S. (2010). Original score. <https://www.flickr.com/photos/stephmouss/5402989572> [CC By 2.0]
- POP. (2011). Dated ms. ownership inscription of the Alsatian humanist Beatus Rhenanus (1485-1547). <https://www.flickr.com/photos/58558794@N07/5400585187> [CC By 2.0]
- TERN. (2014). TERN flux tower site - Tumberumba. <http://fluxnet.ornl.gov/site/43>

- ANDS curation continuum <http://ands.org.au/assets/images/curation.continuum.gif>
- ANDS data citation poster <http://ands.org.au/cite-data/images/data-citation-poster-medium.png>
- 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:ANd9GcRESNMvW44FJQ0x-7VJ_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>
- 2009_03alab notebook http://www.labtimes.org/labtimes/method/methods/img/2009_03a.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).