



Australian Research Data Commons

Data... so what's the problem?

eResearch NZ 2020 – United in Data

PRESENTED BY

Rosie Hicks, CEO, ARDC Ltd

Rosie.Hicks@ardc.edu.au



NCRIS – National Collaborative Research Infrastructure Strategy

~40,000 domestic & international users / year.

\$3.3 billion invested since 2004, attracting > \$1 billion in co-investment

24 active projects, > 200 institutions, > 1900 experts, researchers & facility managers



A screenshot of the National Collaborative Research Infrastructure Strategy (NCRIS) website. The header includes the Australian Government Department of Education and Training logo and the text 'National Innovation and Science Agenda'. The main heading is 'National Collaborative Research Infrastructure Strategy (NCRIS)'. Below this, there are links for 'Business and Industry', 'Education authorities', 'Government agencies', 'Graduates', 'Media', and 'Students'. The main content area describes NCRIS as a national network of world-class research infrastructure projects that support high-quality research that will drive greater innovation in the Australian research sector and the economy more broadly. Projects support strategically important research through which Australian researchers and their international partners can address key national and global challenges. The page also includes a search bar with the text 'I need to find out about:' and a 'Find' button, and links for 'Tips for searching this site' and 'Search documents'. The footer of the page shows 'Higher Education Research'.

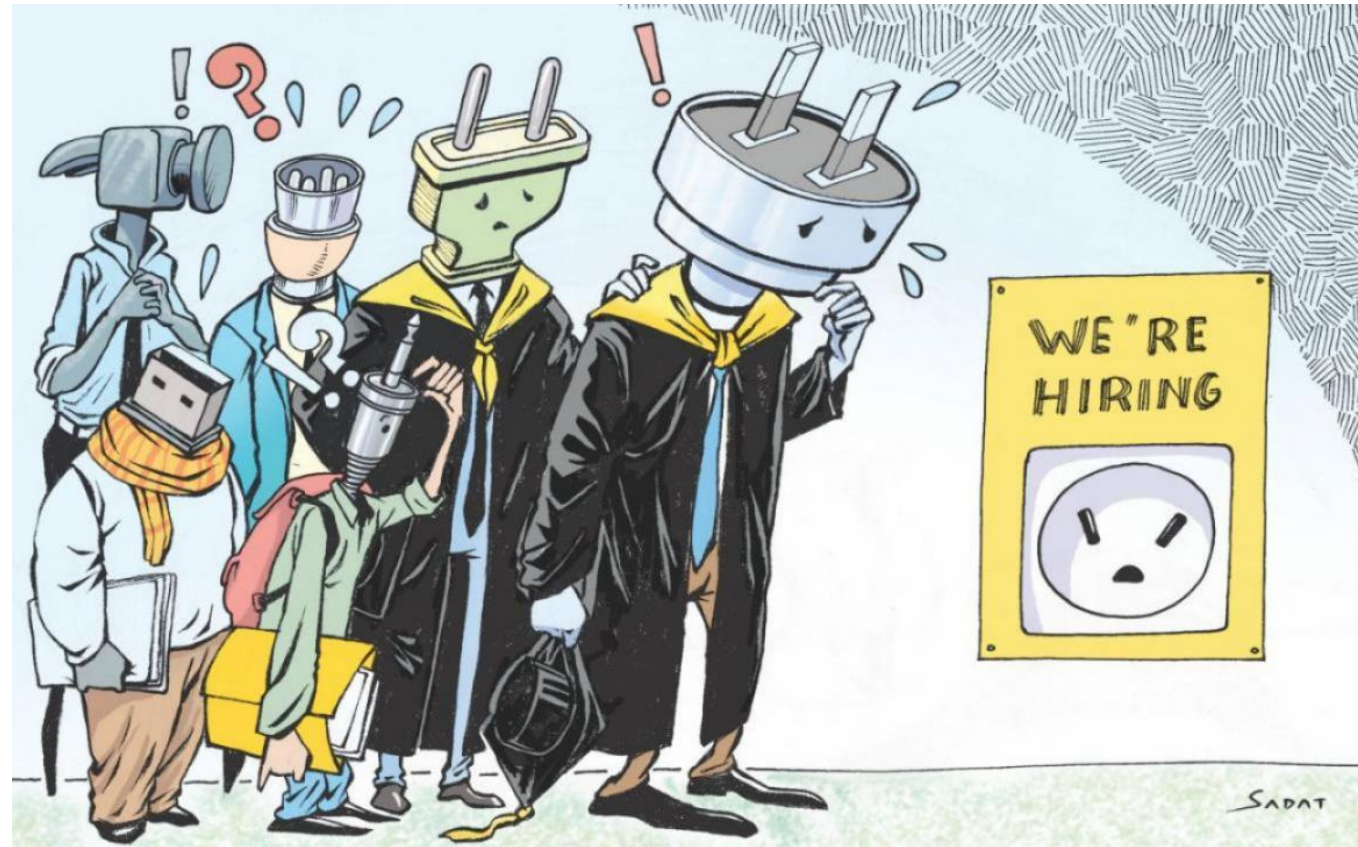
Forming the Australian Research Data Commons



- 2017 - alignment
- 2018 – virtual organisation
- 2019 – incorporation, ARDC Ltd



Challenges: skilled workforce



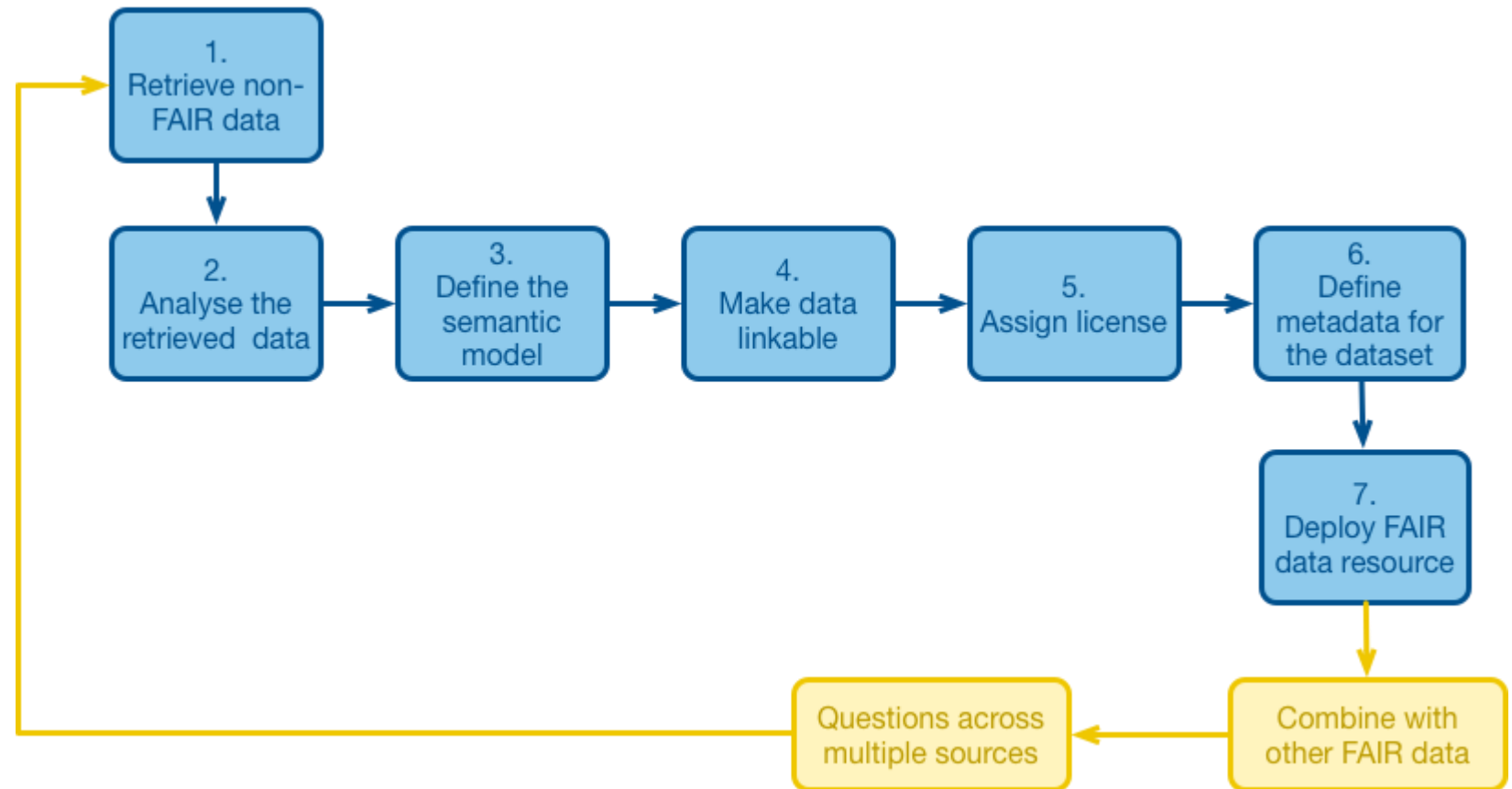
Challenge: Making data more FAIR

F_{indable}

A_{ccessible}

I_{nteroperable}

R_{eusable}



2019 Nature study - >50% not aware of FAIR

■ I am familiar with the FAIR principles ■ I have previously heard of the FAIR principles but I'm not familiar with them ■ I've never heard of the FAIR principles before now

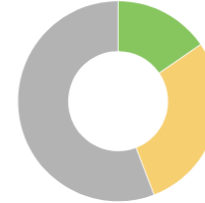
Arts & Humanities



Astron. & Planetary Science



Biology



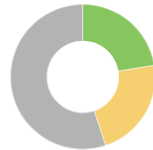
Business



Chemistry



Earth & Env. Science



Engineering



Materials Science



Medicine



Physics



Social Science

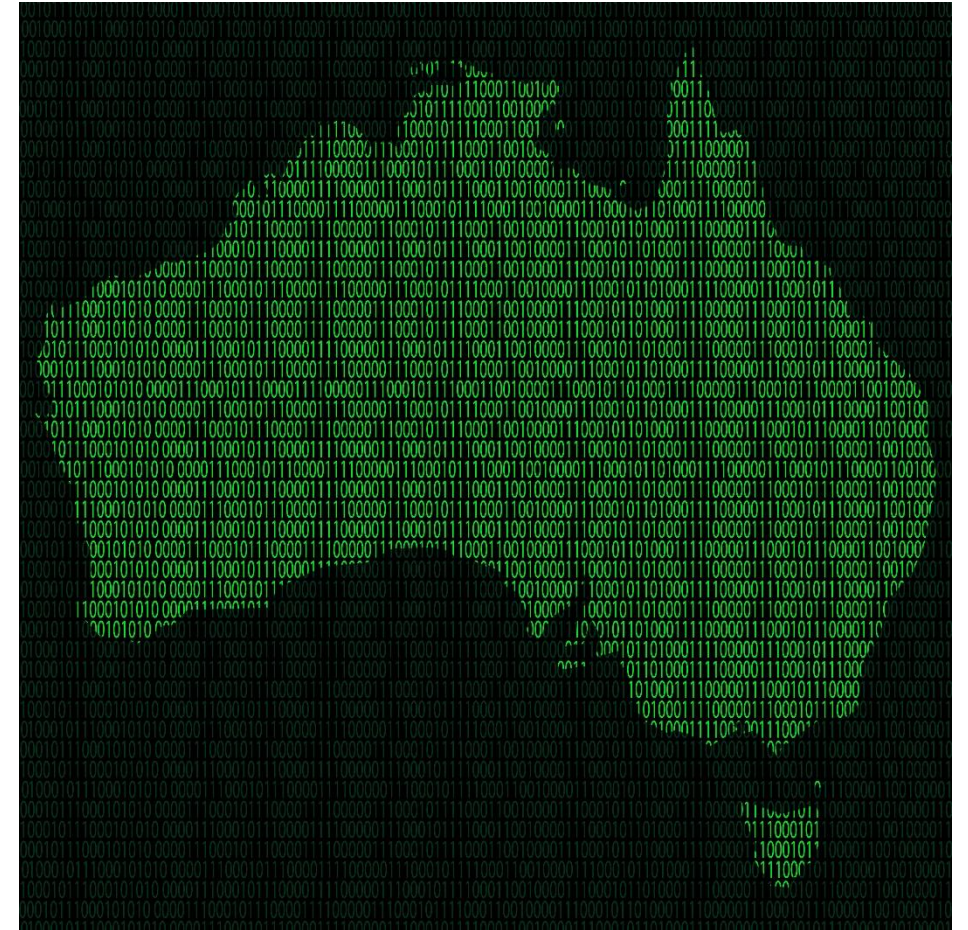
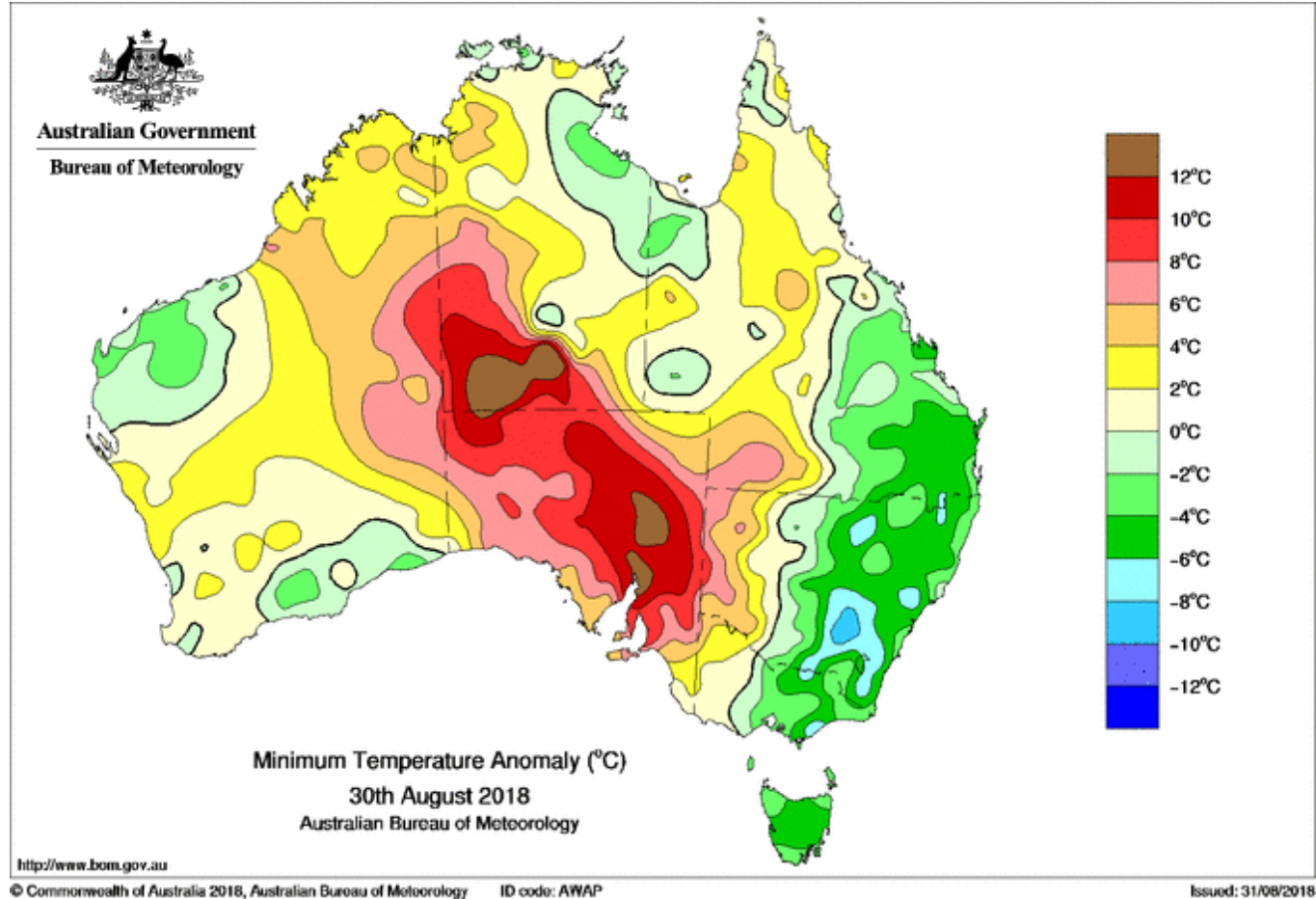


Other



Source: [State of Open Data](#)

Challenge: Access to non-research data



NASA levels of instrument data processing*

Data level	Description
0	Reconstructed unprocessed instrument data at full resolution.
1A	Reconstructed, unprocessed instrument data at full resolution, time referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (i.e., platform ephemeris) computed and appended, but not applied to the Level 0 data.
1B	Level 1A data that has been processed to sensor units (i.e., radar backscatter cross section, brightness temperature, etc.). Not all instruments will have a Level 1B equivalent.
2	Derived environmental variables (e.g., ocean wave height, soil moisture, ice concentration) at the same resolution and location as the Level 1 source data.
3	Variables mapped on uniform space-time grid scales, usually with some completeness and consistency properties (e.g., missing points interpolated, complete regions mosaicked together from multiple orbits).
4	Model output or results from analyses of lower-level data (i.e., variables that were not measured by the instruments but instead are derived from these measurements).

Challenge: Data Quality

CENSUS OF IRELAND, 1901.
(Two Examples of the mode of filling up this Table are given on the other side.)
FORM A.
No. on Form B. 62
of this FAMILY and their VISITORS, BOARDERS, SERVANTS, &c., who slept or abode in this House on the night of SUNDAY, the 31st of MARCH, 1901.

RELATIVE TO HEAD OF FAMILY	RELIGIOUS PROFESSION	EDUCATION	AGE	SEX	RANK, PROFESSION, OR OCCUPATION	MARRIAGE	WHERE BORN	IRISH LANGUAGE	If Dead and Dumb: Blind, Imbecile or Idiot; or Lunatic.
1. <i>Thomas O'Donnell</i>	<i>Roman Catholic</i>	<i>Read & write</i>	<i>23</i>	<i>M</i>	<i>Publican</i>	<i>Married</i>	<i>Ireland, Co. Wick</i>	<i>Write the word "Irish" in this column opposite the name of each person who speaks Irish only, and the words "Irish & English" opposite the names of those who can speak both languages. In both languages, in other cases no entry should be made in this column.</i>	
2. <i>William O'Donnell</i>	<i>Roman Catholic</i>	<i>Read & write</i>	<i>20</i>	<i>M</i>					
3. <i>Samuel O'Donnell</i>	<i>Roman Catholic</i>	<i>Read & write</i>	<i>17</i>	<i>M</i>					
4. <i>Ann O'Donnell</i>	<i>Roman Catholic</i>	<i>Read & write</i>	<i>16</i>	<i>F</i>					

I hereby certify, as required by the Act 63 Vic, cap 6, s 6 (1), that the foregoing Return is correct, according to the best of my knowledge and belief.
Peter O'Donnell (Signature of Head of Family)

CENSUS OF IRELAND, 1911.
(Two Examples of the mode of filling up this Table are given on the other side.)
FORM A.
No. on Form B. 134
of this FAMILY and their VISITORS, BOARDERS, SERVANTS, &c., who slept or abode in this House on the night of SUNDAY, the 2nd of APRIL, 1911.

RELATIVE TO HEAD OF FAMILY	RELIGIOUS PROFESSION	EDUCATION	AGE (not Birthdate) and SEX	RANK, PROFESSION, OR OCCUPATION	PARTICULARS AS TO MARRIAGE	WHERE BORN	IRISH LANGUAGE	If Dead and Dumb: Blind, Imbecile or Idiot; or Lunatic.
1. <i>Thomas O'Donnell</i>	<i>Roman Catholic</i>	<i>Read & write</i>	<i>30</i>	<i>Publican</i>	<i>Married</i>	<i>Ireland, Co. Wick</i>	<i>Write the word "Irish" in this column opposite the name of each person who speaks Irish only, and the words "Irish & English" opposite the names of those who can speak both languages. In both languages, in other cases no entry should be made in this column.</i>	
2. <i>William O'Donnell</i>	<i>Roman Catholic</i>	<i>Read & write</i>	<i>26</i>		<i>Married</i>			
3. <i>Samuel O'Donnell</i>	<i>Roman Catholic</i>	<i>Read & write</i>	<i>23</i>		<i>Married</i>			
4. <i>Ann O'Donnell</i>	<i>Roman Catholic</i>	<i>Read & write</i>	<i>20</i>		<i>Married</i>			

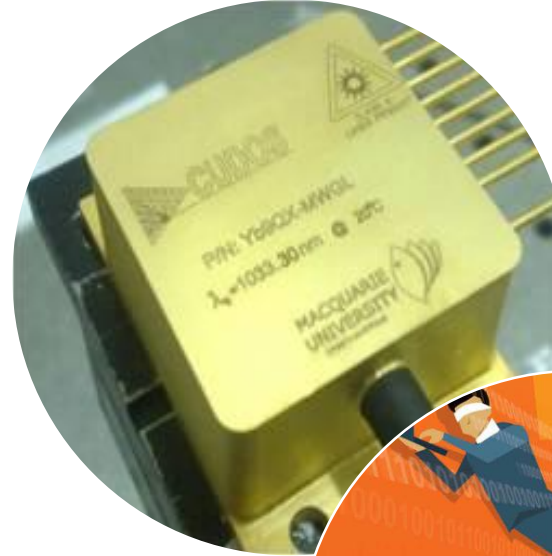
I hereby certify, as required by the Act 10 Edw. VII, and 1 Geo. V, cap. 11, that the foregoing Return is correct, according to the best of my knowledge and belief.
John O'Donnell (Signature of Head of Family)

Challenge: Trusted data

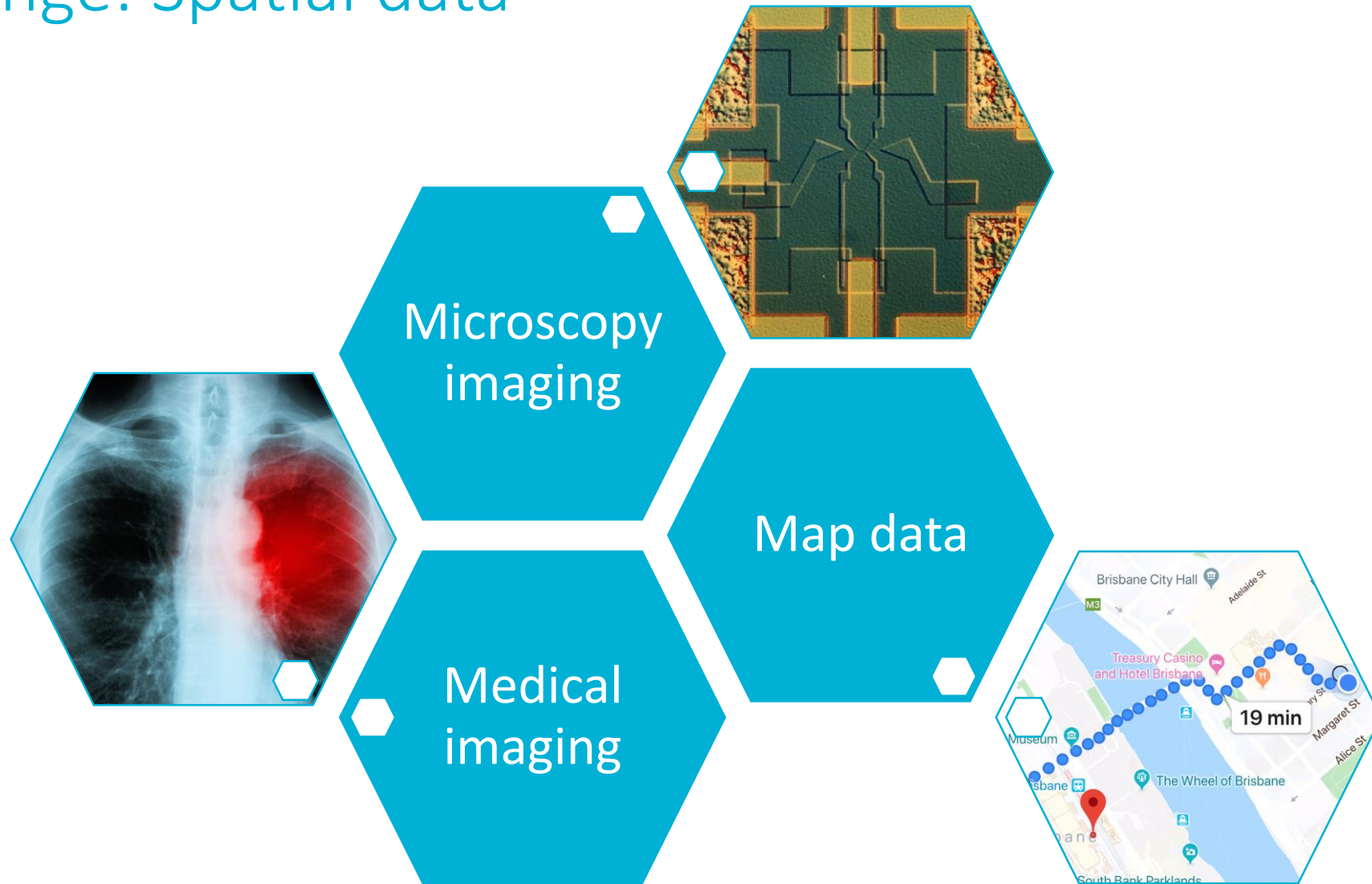


- Data repository core level certification based on the [DSA-WDS Core Trustworthy Data Repositories Requirements](#)
- Trusted Repositories Audit and Certification (TRAC)

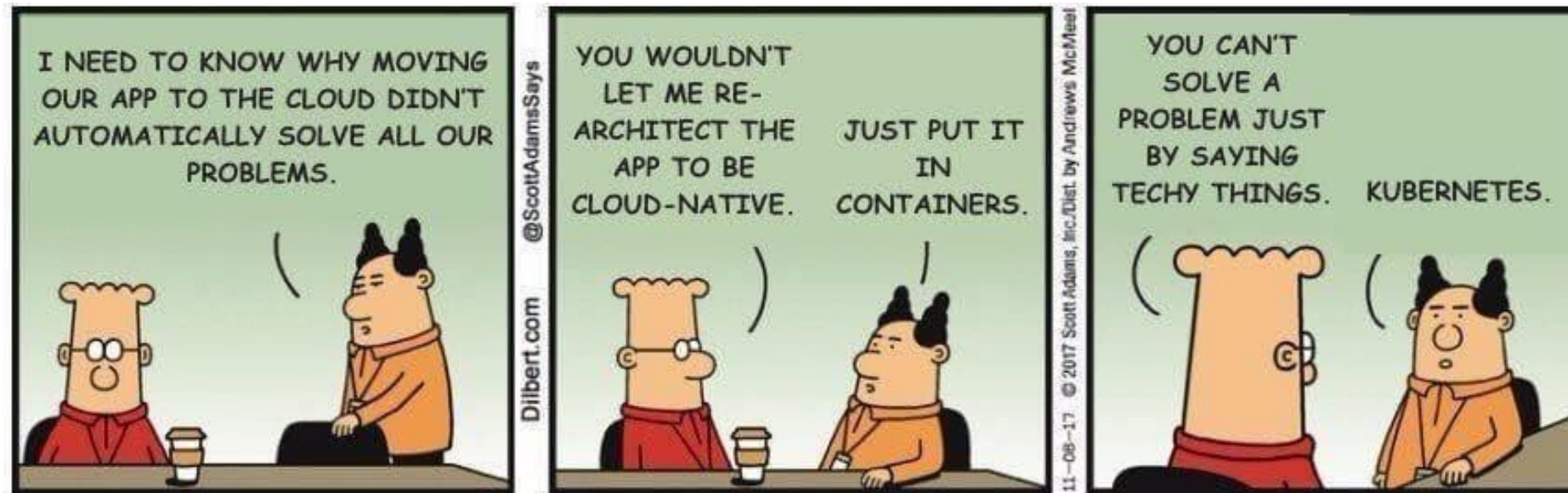
Challenge: Sensitive data



Challenge: Spatial data



Technologies...



Technologies: AI

Machine learning

Robotics

Facial recognition

Voice assistants



AFR 100 Women of Influence 2019



Australian Research Data Commons

ARDC \neq *The Australian Research Data Commons*

Together with partners, we will build the commons



Australian Research Data Commons

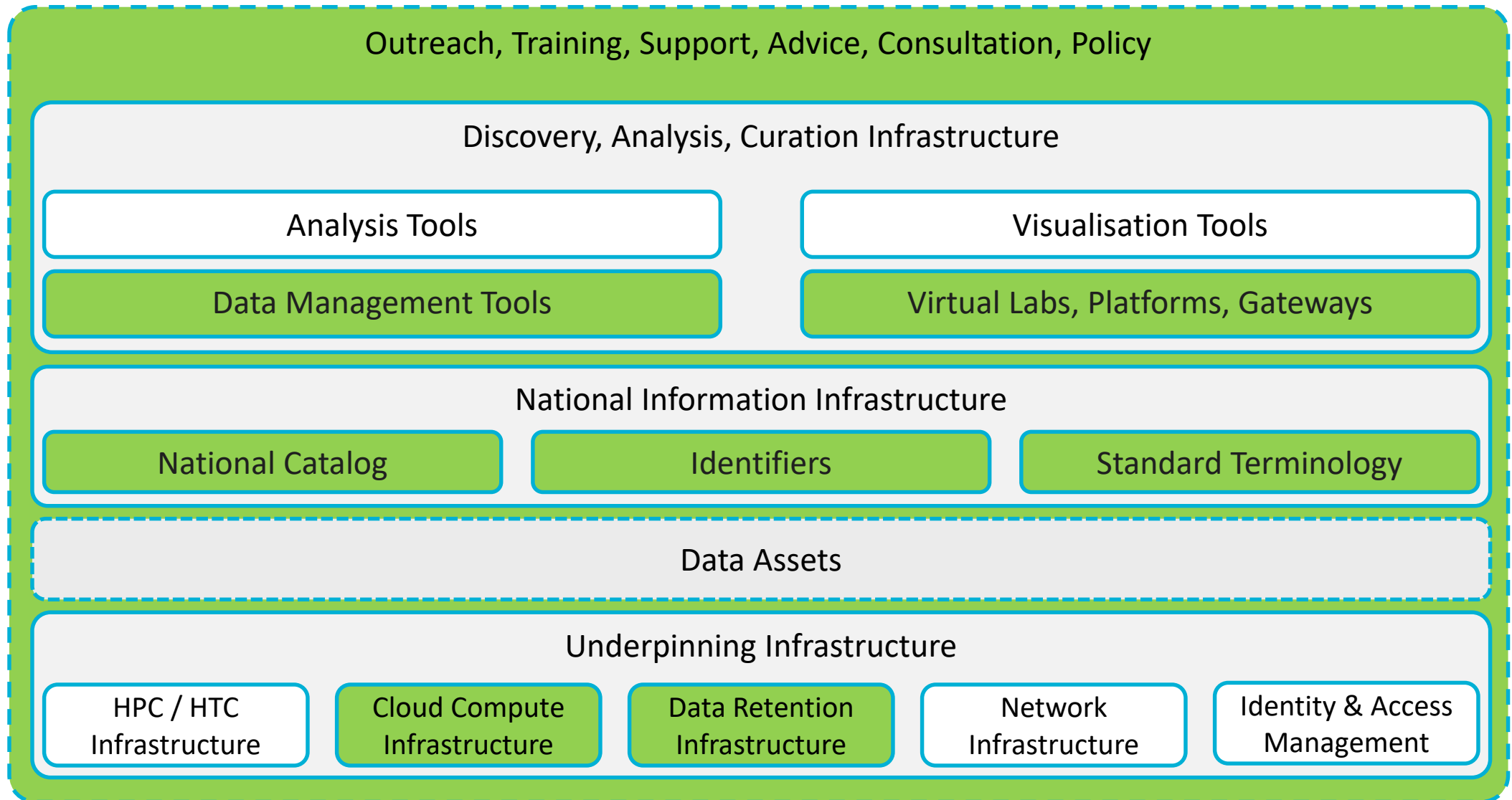
ARDC in Context: Digital Data and eResearch Platforms

This national eResearch infrastructure area is a cross-cutting capability that serves research collaboration, modelling, data and data analysis needs. It comprises advanced networks [AARNet]; identity, access and authentication services [AAF]; high performance and cloud computing resources [NCI, Pawsey]; management of and access to research data; the development and adoption of new digital research techniques; and the integration of all those elements to create digital environments researchers use every day [ARDC]. Research increasingly depends on digital evidence and related data and on digital methods as a new means to progress ideas and advance knowledge. As such, the ability to support those activities through more effective digital data and eResearch platforms becomes critical

(2016 National Research Infrastructure Roadmap)

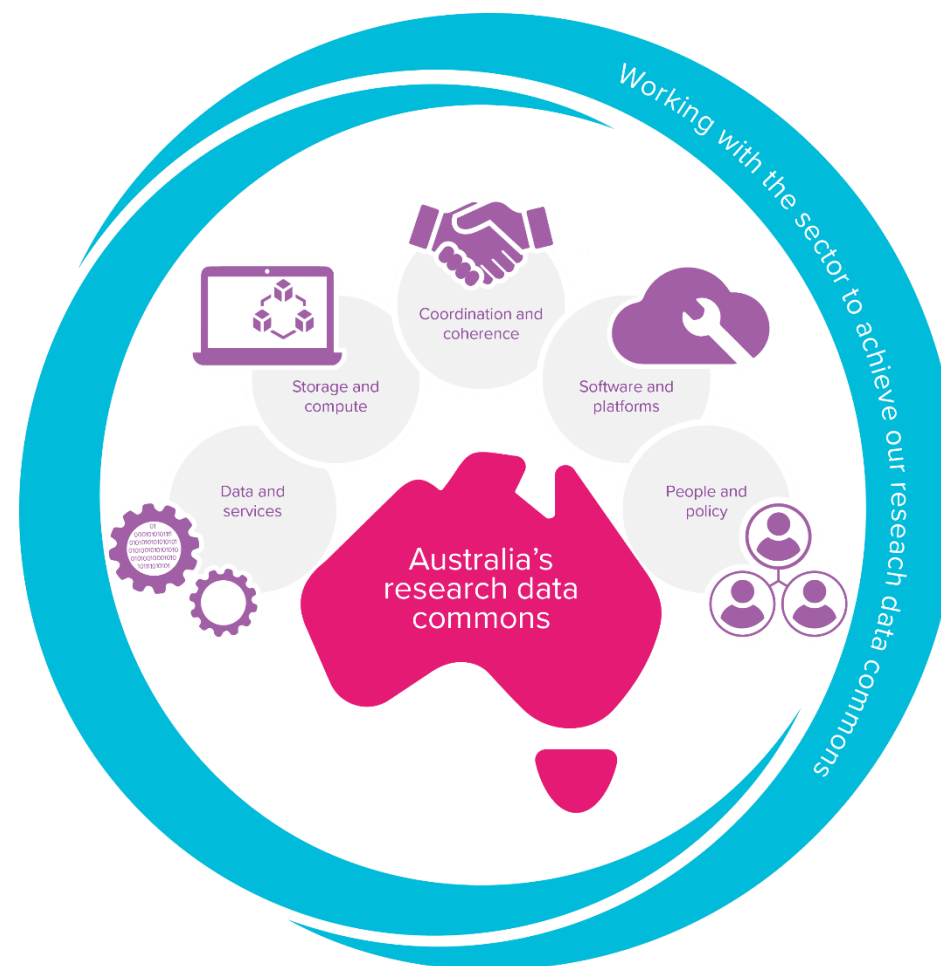


ARDC's areas of activity



ARDC – The Australian Research Data Commons

ARDC supports research across all disciplines: to enable world leading research, facilitate accelerated innovation, and enhance researchers' ability to translate outputs into societal benefit



ARDC – Five Themes

Coordination and Coherence

Facilitating an “Australian research data commons”

People and Policy

Transforming culture and community

Data and services

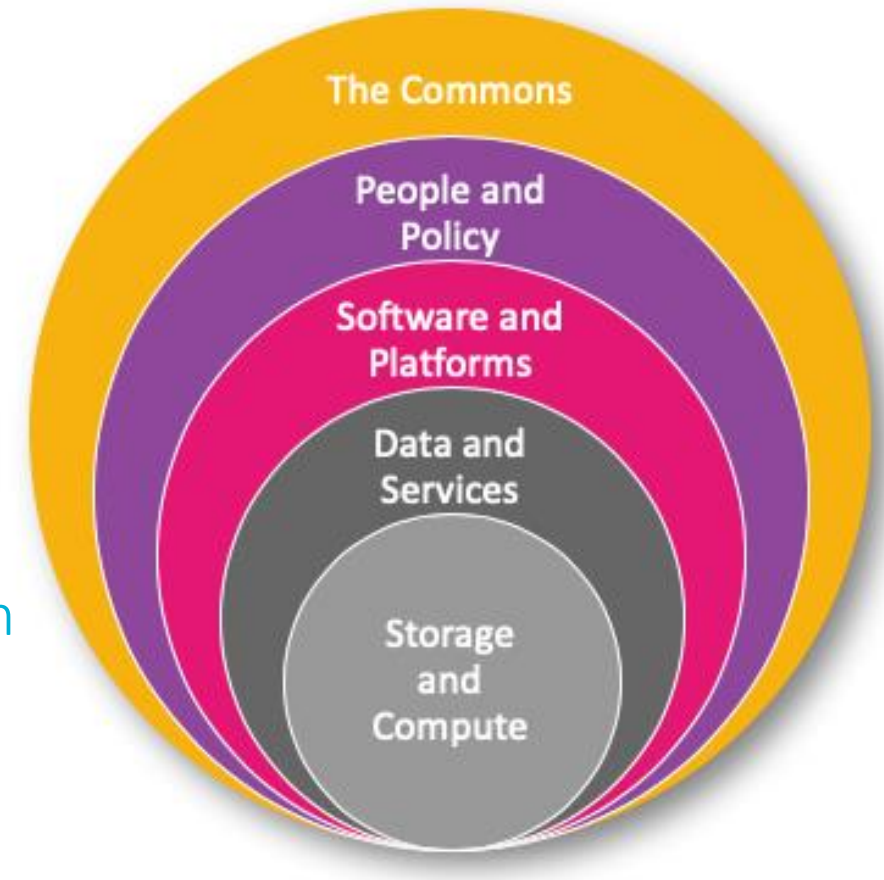
Maximising the value of Australia’s data assets

Software and Platforms

Enabling research insights & supporting collaboration

Storage and Compute

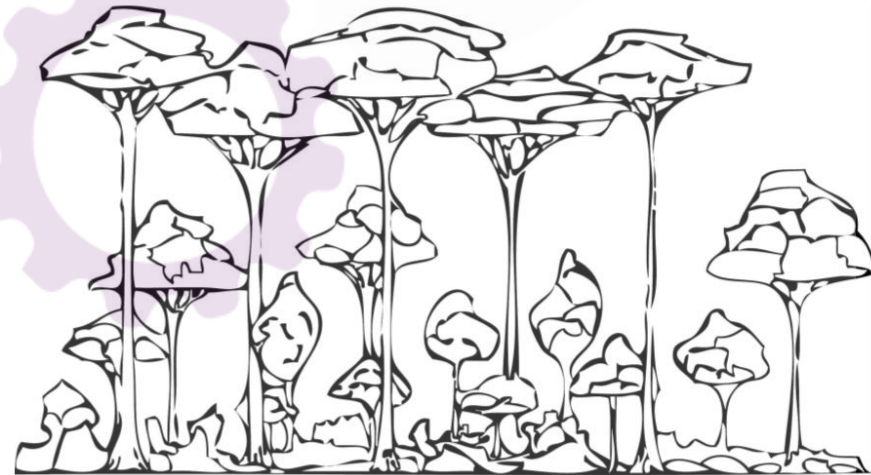
Providing foundation infrastructure



Data and Services: Maximising the value of Australia's data assets

- Reference data collections
- Sensitive data and approaches, platforms and services to manage, collaborate over and share this data
- Data archives and trusted data repositories
- Ongoing support for [Research Data Australia](#) and associated data publishing services (Research Vocabularies Australia, DataCite DOI, etc)

National Data Rainforest



- Reference Collections

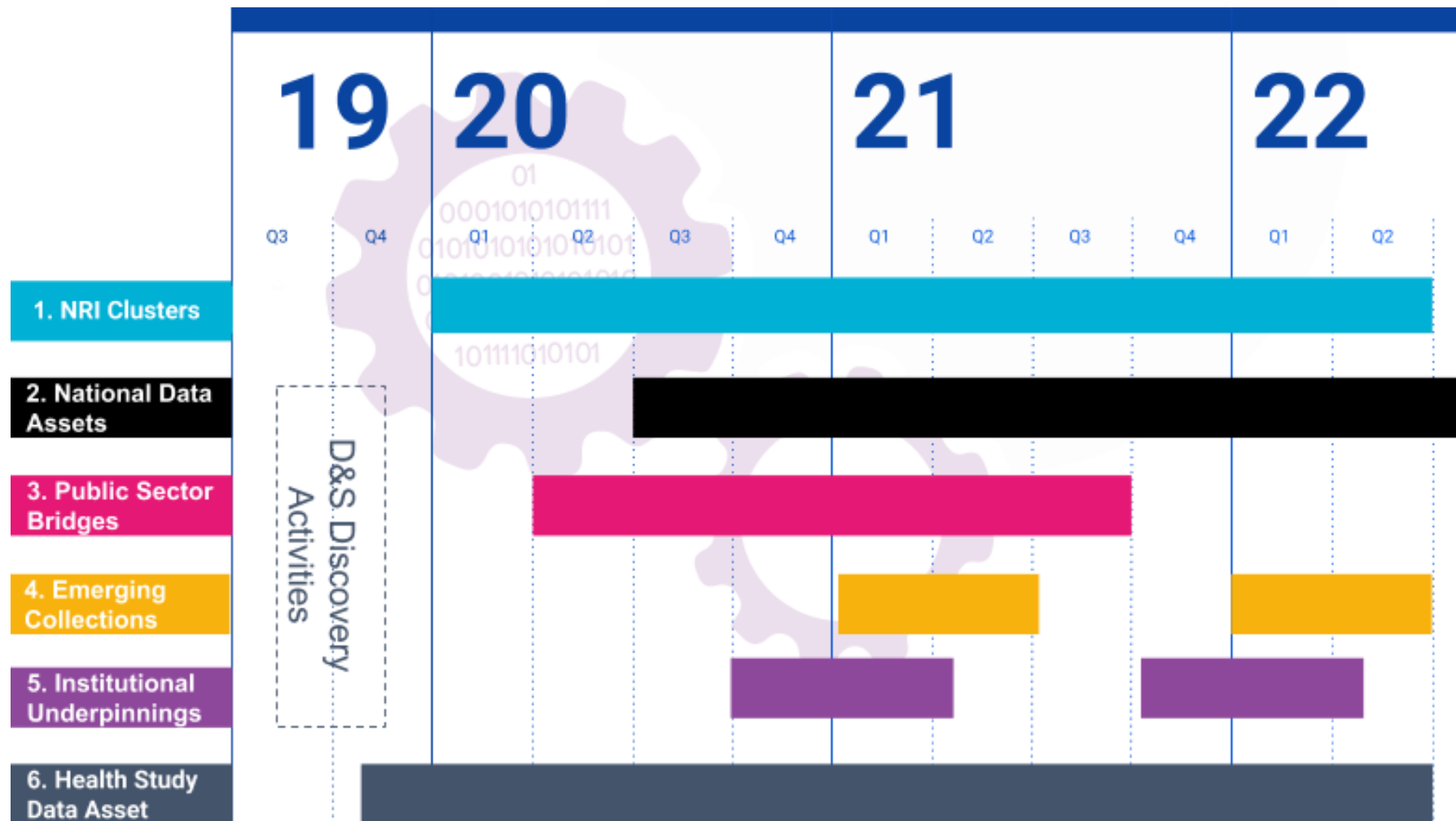
- Community Collections

Data for Efficiency, Integrity, Reproducibility.

Research Institutions, Govt Agencies, Utilities, National Facilities....



Data and Services: Maximising the value of Australia's data assets

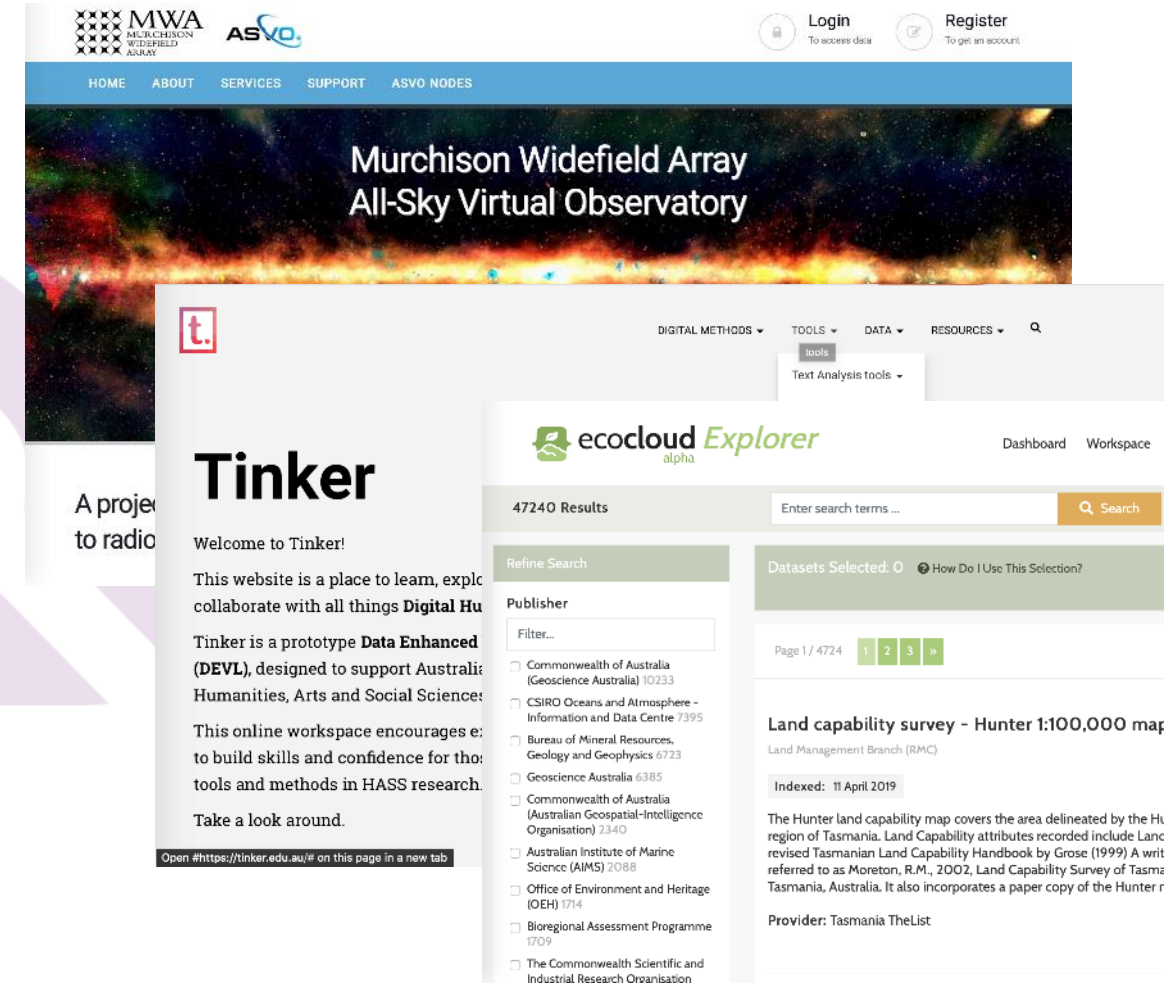


Some examples...

Program	Project	Lead
NRI Clusters	Developing the Water and Energy Supply and Consumption Data Standard	AURIN
National data assets	Enactment of the Agricultural Research Data Federation (AgReFed) Data and Stewardship and Governance Framework and the formation of AgReFed	Federation Uni
Public sector bridges	Preserving the Australian Census - 250 years of population data for Australia	ADA
Emerging collections	Overcoming pinch-points in ingesting, cataloguing and accessing (meta)data for the development of a national language data commons	Uni Queensland
Institutional underpinnings	FAIR Simple Scalable Static Research Data Repository Demonstrator	UTS
Health Studies Data Asset	What large clinical data sets exist across the member organisations of MACH, Melbourne and how can we collate and curate them to maximise research outcomes?	Melbourne Uni

Software and Platforms: Supporting collaborative research

- Deliver world-leading informatics capabilities
- Priorities:
 - High-quality data workflows
 - Analysis capabilities
 - Data to Modelling
- Areas of initial focus: VLs (or platforms..)
 - What does VL V.2 look like?
 - Build on existing VLs
 - Open up for new VLs





ecocloud: EcoScience Research Data Cloud

What does this project enable?

- Access to more trusted data
- Species trait modelling analysis tools
- Data portal with access to climate and Essential Environmental measures data
- Training to ensure effective uptake

What is the impact?

- Analyses and models of the GlobalArchive fish data in ecocloud supports National State of the Environment reporting and adds value to the millions of dollars already invested in the dataset



Characterisation Data Enhanced Virtual Laboratory

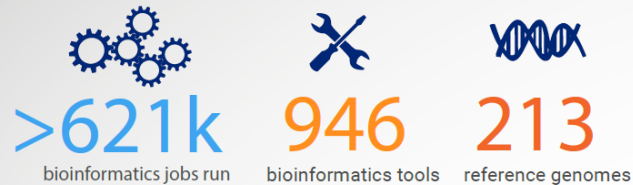
What does this project enable?

- Rich online environments for characterisation in the cloud and on HPC platforms for specific areas
- Development of a national network of characterisation informatics experts
- Production of a national training network and program of work to uplift data skills across characterisation users

What is the impact?

- Researchers are creating detailed images of an important class of cell surface receptors, with the prospect of identifying new and/or better drugs. Sexton (NHMRC Senior Principal Research Fellow) & colleagues are studying ~20 different types of G protein-coupled receptors and the roles they play in diseases such as **diabetes, obesity and cardiovascular disease**

A complete bioinformatics analysis platform with:



... and on average 2 new tools or references added weekly on user request

Supporting published research across many areas:



54
publications mentioned the GVL

31
studies used the GVL in their publication

An active and engaged user community

4545
registered users.
724 active users (last 90 days)

User growth 2016 - 2019 Q2



Registered users in Australia from:

30
Australian Universities

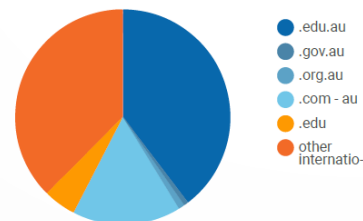
22
Medical Research Institutes or Organisations

19
Other Research Organisations

4
teaching courses utilising GVL annually

1064
people trained in hands-on workshops

Users per domain



Users represented across
420
organisations

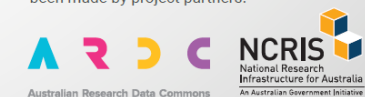
Users represented across
49
countries



"We have sequenced well over 500 transcriptomes/genomes, and routinely use Galaxy Australia, powered by the GVL, for many bioinformatics processes. It is easy to use, has high computational power, a sophisticated support structure and enables global collaboration through straightforward data sharing. We greatly appreciate the service."

- Dr Fabio Cortesi and Prof Justin Marshall, Queensland Brain Institute
 Science Advances 2017 3(11), eaao4709

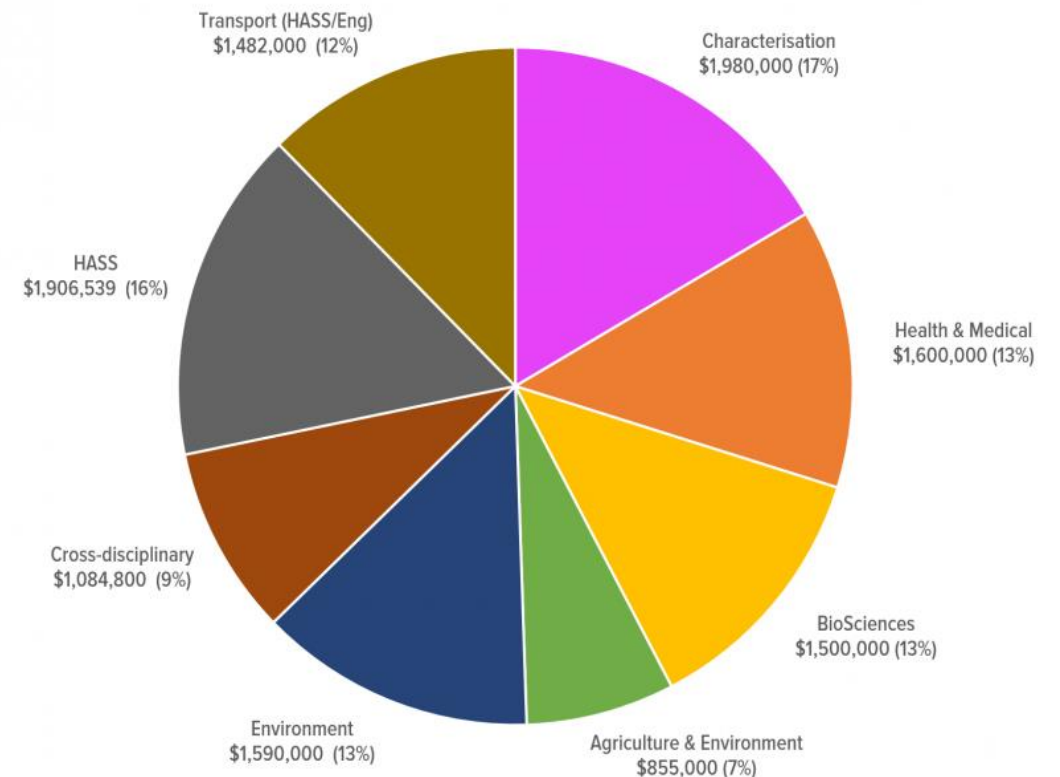
Funded by the Commonwealth Government National Collaborative Research Infrastructure Strategy (NCRIS). Co-investments have also been made by project partners.



All statistics are Jan 2016 - Aug 2019

2019 Platforms Call – successful projects

- Environments to Accelerate Machine Learning Based Discovery
- E-Research Institutional Cloud Architecture (ERICA)
- BioCommons Bring Your Own Data (BYOD)




Storage and Compute: Providing foundation infrastructure

- Reward the provision of FAIR data, tools & resources while supporting infrastructure agility
 - impact metrics to be part of the “reward” mechanism in the longer term
- Reflect who is *responsible* for FAIR assets
 - To support sustainable models
- Foster a constructively competitive underpinning infrastructure environment
- Areas of initial focus:
 - “consumption” based funding
 - avoid capital cliffs
 - potential challenges around co-investment
 - Short, medium, and long-term priorities



People and Policy: Transforming culture and community

- Policy
- Workforce development, Skills Program
- Collaborations/Partnerships
 - Communities (bringing together participants)
 - Skills community broker
- Areas of initial focus
 - Initial scoping and mini-projects
 - Skills summit

A banner image for ARDC events. It shows a person in a red jacket standing at a podium on a stage, with the word 'Events' overlaid in white text. The ARDC logo is at the top. The background is dark with vertical light stripes.

Events

The Australian eResearch Skilled Workforce Summit

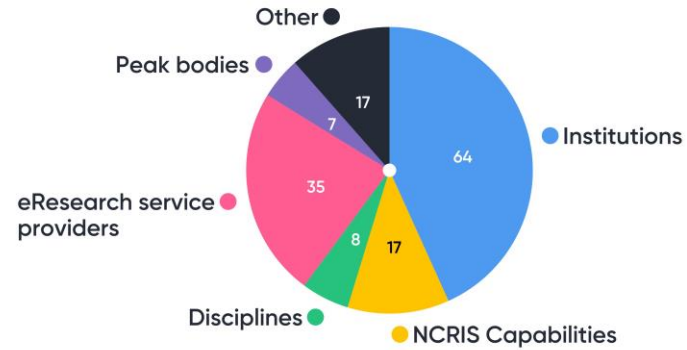
29 - 30 Jul 2019
Sydney

The Australian eResearch Skilled Workforce Summit will take place on 29 and 30 July 2019 in Sydney. It is being coordinated by the Australian Research Data Commons (ARDC), Council of Australasian University Directors of Information Technology (CAUDIT) and Universities Australia (UA).

The Summit will focus on the development of the digitally skilled workforce, essential for research in the 21st century. It

Who is in the audience? (can choose more than one option)

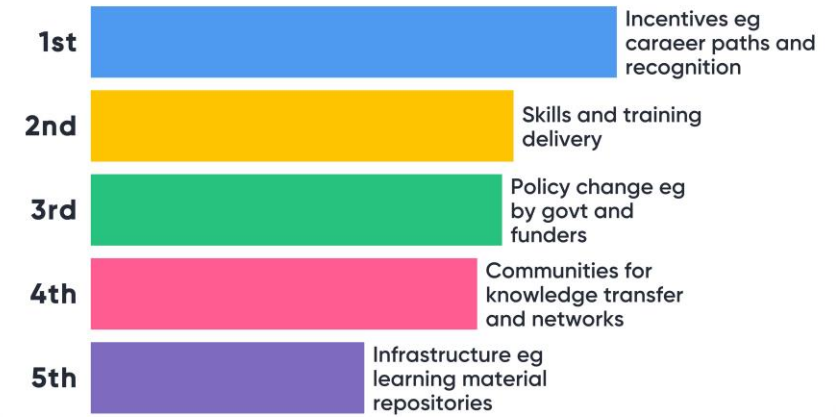
Mentimeter



107

Rank these 5 items needed for cultural change in terms of where effort is most needed now.

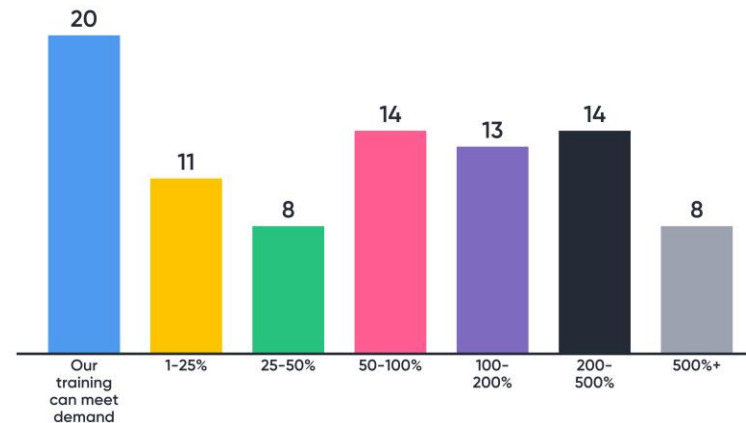
Mentimeter



79

With regard to training initiatives that you're involved with, by how much does demand outstrip supply?

Mentimeter



79

Skilled Workforce Summit – key outcomes

1 – Identified the sector's priority areas of concern

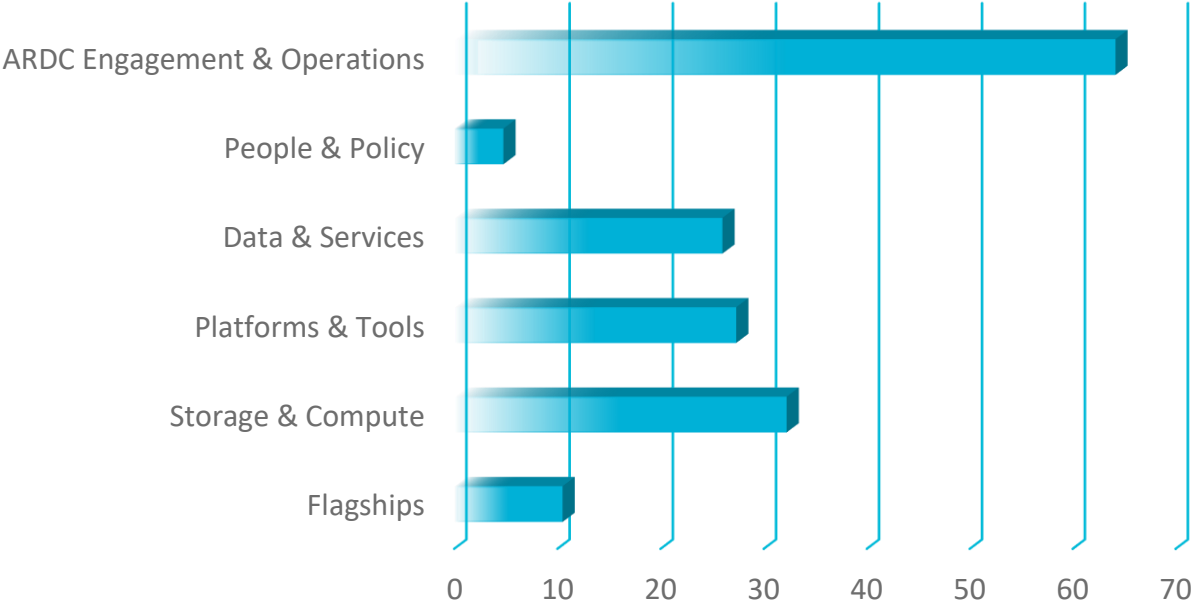
#2 - Reinforced the need for the national coordination by ARDC

Goal: to ensure researchers have the awareness, skills and support to realise the benefits of high quality data assets and infrastructure



The details...

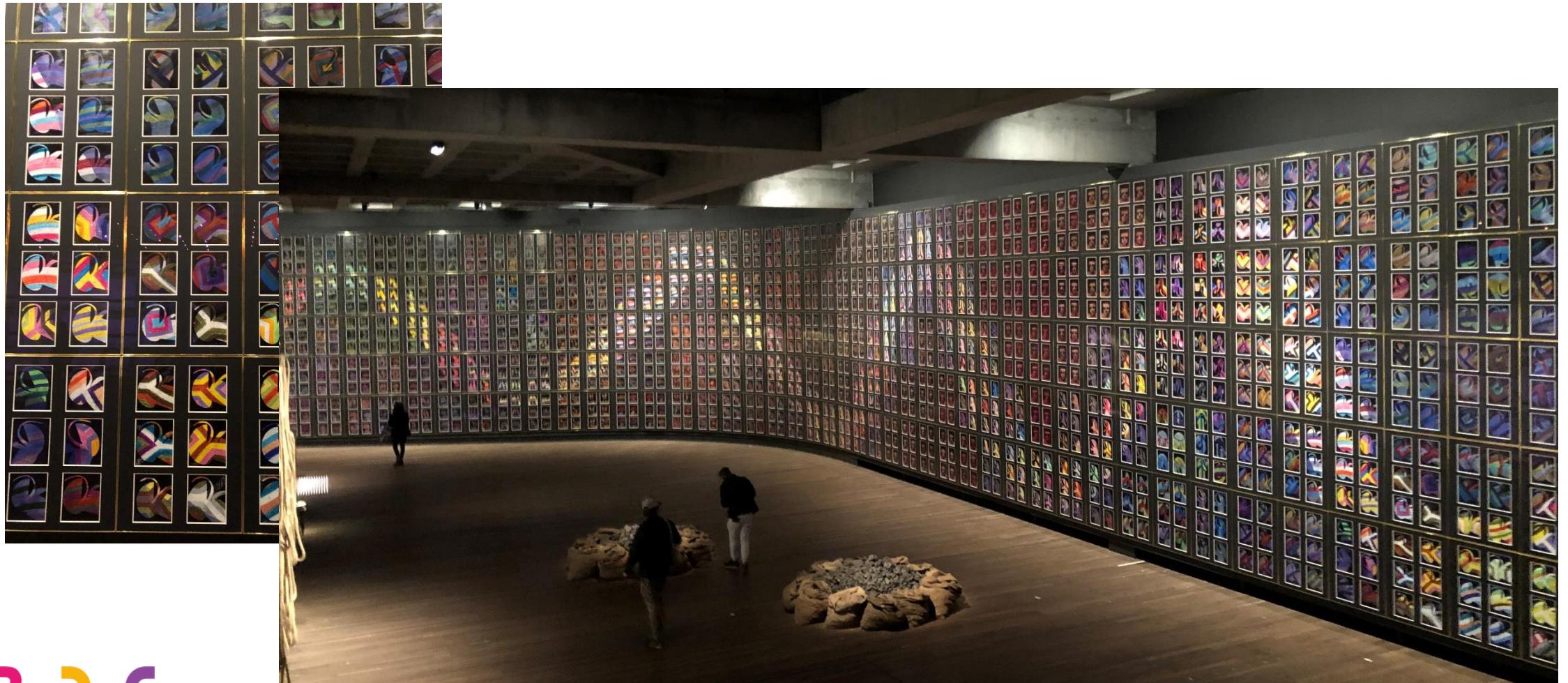
ARDC BUDGET (AU\$ MILLIONS)

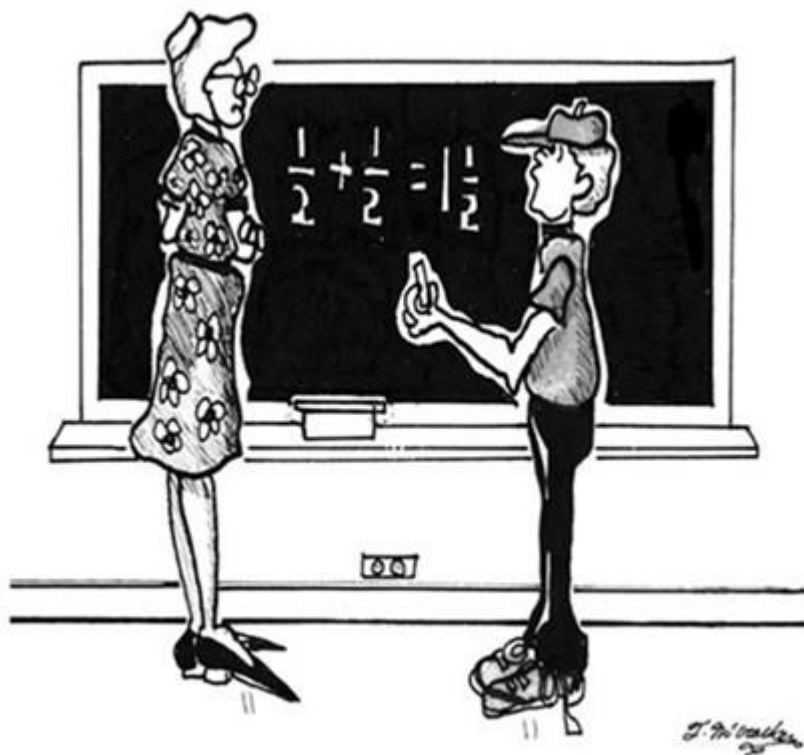


Coordination and coherence



The role of ARDC





"BUT DIDN'T YOU ONCE TELL US THAT, 'THE WHOLE
IS GREATER THAN THE SUM OF ITS PARTS?'"

ARDC Staff retreat, April 2019

NCRIS
National Research
Infrastructure for Australia



Australian Research Data Commons



Theme: Data for real world impact

rd-alliance.org

@resdataall #RDAPlenary



Australian Research Data Commons



Australian Research Data Commons

*Transforming digital infrastructure to support
leading edge research and innovation*



Australian Research Data Commons

