



Australian Research Data Commons

# Trends in Research, Research Data and Responses

PRESENTED BY

Ross Wilkinson  
Australian Research Data Commons  
Canberra, 2019





## Outline:

- Some research trends
- The Australian research data experience
- Some reflections

## Some Research Trends:

- Scale of Problem
- Complexity
- Translation of research for society benefit
- Reproducible research
- High reliability research
- Open research
- Leave it to AI!

# The challenge to Oldenburg's principle: - a crisis of replicability and credibility?

NATURE | VOL 483 | 29 MARCH 2012

## REPRODUCIBILITY OF RESEARCH FINDINGS

Preclinical research generates many secondary publications, even when results cannot be reproduced.

Journal impact factor	Number of articles	Mean number of citations of non-reproduced articles*	Mean number of citations of reproduced articles
>20	21	248 (range 3–800)	231 (range 82–519)
5–19	32	169 (range 6–1,909)	13 (range 3–24)

Results from ten-year retrospective analysis of experiments performed prospectively. The term 'non-reproduced' was assigned on the basis of findings not being sufficiently robust to drive a drug-development programme.

\*Source of citations: Google Scholar, May 2011.

**A fundamental principle: the data providing the evidence for a published concept MUST be concurrently published, together with the metadata**

**But what about the vast data volumes that are not used to support publication as well as those that are?**



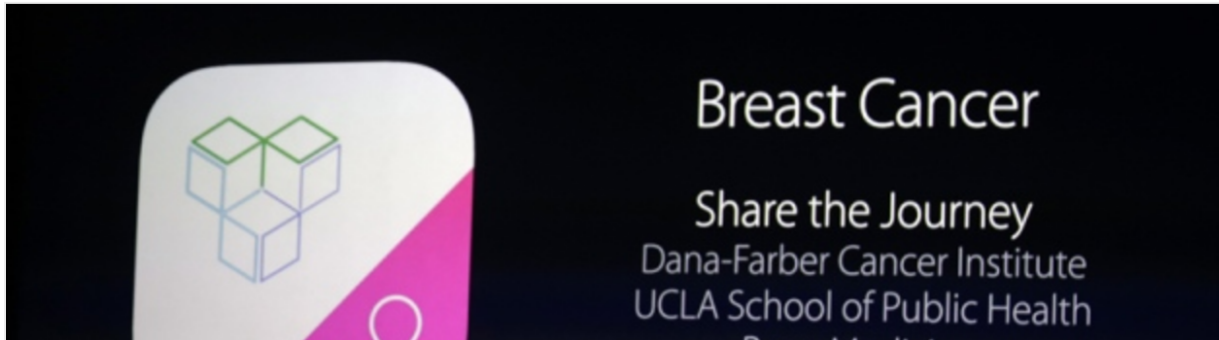
# Smartphones set to boost large-scale health studies

Tech giant Apple introduces mobile platform for biomedical research.

Helen Shen

10 March 2015

Rights & Permissions



## Brain development



### How poverty affects the brain

An unprecedented study in Bangladesh could reveal how malnutrition, poor sanitation and other challenges make their mark on child development.

Recent Read Commented

1. [Engineered cell therapy for cancer gets thumbs up from FDA advisers](#)  
Nature | 12 July 2017
2. [Information technology: A digital genius at play](#)

- Clinical trial with 100K+ participation
- Data is not “controlled”



## **Equitable Access to Sustainable Development**

- secure and resilient food systems supported by sustainable marine resources and agriculture
- sustainable health and well being
- inclusive and equitable quality education
- clean air, water and sanitation
- affordable, reliable, sustainable energy

## **Sustainable Economies and Societies**

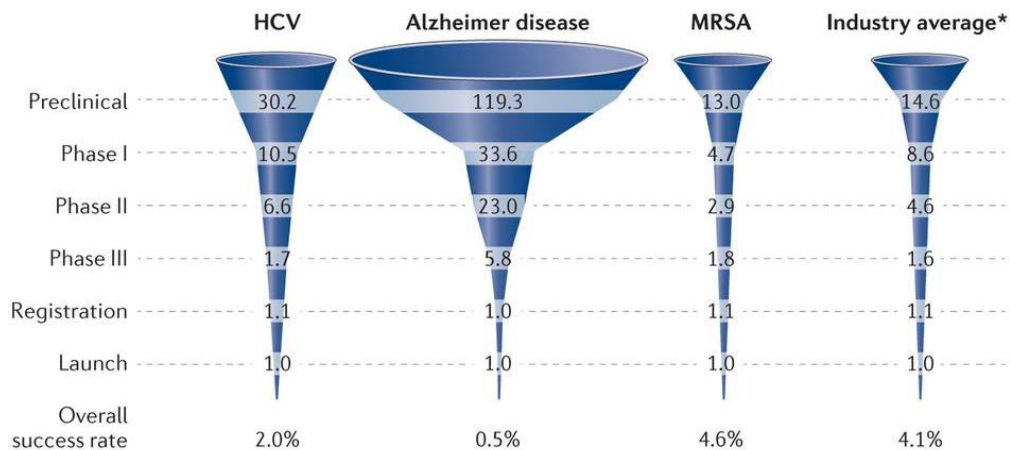
- Sustainable livelihoods supported by strong foundations for inclusive economic growth and innovation

- Resilience and action on short-term environmental shocks and long-term environmental change
- Sustainable cities and communities
- Sustainable production and consumption of materials and other resources

## **Human Rights, Good Governance and Social Justice**

- Understand and respond effectively to forced displacement and multiple refugee crises
- Reduce conflict and promote peace, justice and humanitarian action
- Reduce poverty and inequality, including gender inequalities.

# Productivity & Irreproducibility

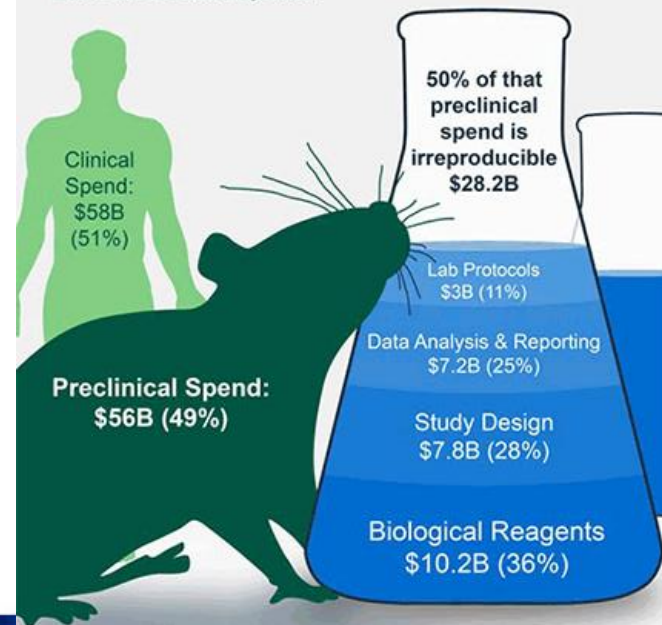


Nature Reviews | Drug Discovery

Paul et al. (Nature Rev. Drug Discov. 9, 203214; 2010)  
Calcoen D, Elias L, Yu X. (Nature Rev. Drug Discov. 14. 161-2; 2015)

## The Economics of Irreproducibility

In the U.S., we spend \$114 billion annually on life sciences research & development. Let's trace that spend.







## AI will solve it

- Recent studies of US sentencing shows systemic increased sentencing of African Americans
- Poor people pay more for photocopy paper in New York
- Best facial recognition company is Chinese
- Google translate is highly usable
- ***The data matters***



## Data Trends:

- FAIR Data
- Open Data
- Data Citation
- Publisher data requirements
- Data Journals
- Funded Fair Data
- Data for Translation
- Trusted data
- Data Repositories
- **Data Quality**
- **Trusted Data Repositories**
- **The Data Marketplace**



# FAIR data for Open Science

- Findable – data should be published with a persistent identifier
- Accessible – it should have an open license – as open as possible
- Interoperable – use community agreed formats, language and vocabularies
- Reusable – rich provenance, enabling use beyond the original purpose



# Data Connectivity (International)

- Research data identification - DataCite
- Researcher Identification – ORCID
- Research publication identification
- Research project identification - RAID
- Research Institution identification
- Research Funder Identification
- Vendor support



**DataCite**  
FIND, ACCESS, AND REUSE DATA



ORCID

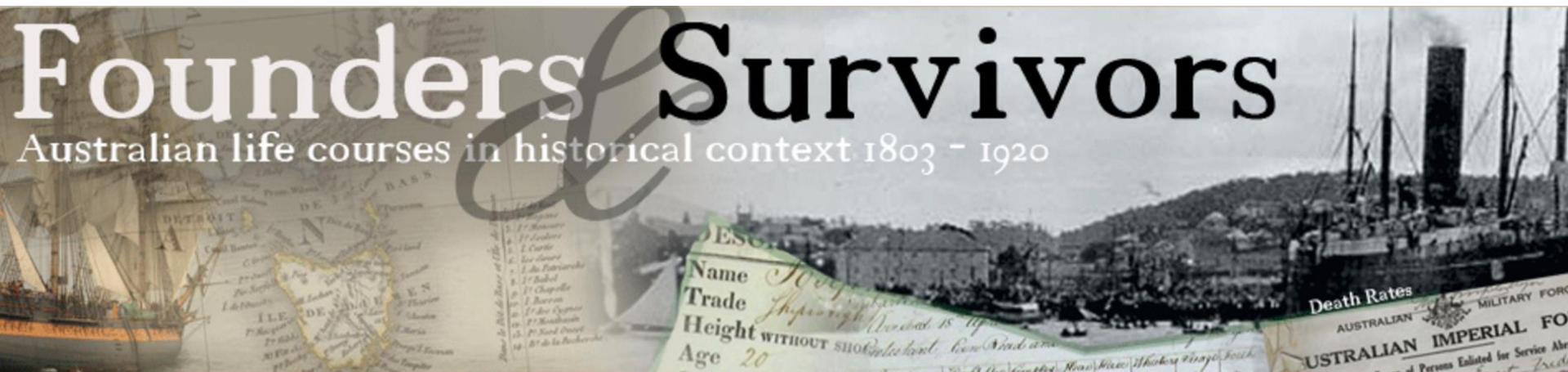
Connecting Research  
and Researchers



## Some International Responses:

- DataCite, Orcid, etc
- Data Journals
- Harmonisation of data policies by funders
- <http://www.nature.com/sdata/policies/repositories>
- International domain initiatives
- Open Data Initiative
- Research Data Alliance
- European Open Science Cloud

Data Complexity – integrating many data sources, from many partners



# Founders & Survivors

Australian life courses in historical context 1803 - 1920

- [Home](#)
- [Search convicts](#)
- [Search site](#)
- [Become involved](#)
- [Forums](#)
- [About Us](#)
- [FAQ](#)
- [Contact us](#)
- [My account](#)
- [Research](#)

[Search for a Convict](#)

[User links](#)

**Founders & Survivors is a partnership between historians, genealogists, demographers and population health researchers. It seeks to record and study the founding population of 73,000 men women and children who were transported to Tasmania. Many survived their convict experience and went on to help build a new society.**

# Data Reliability for Research

- Good news!
- Increasing QA processes from labs generating data for research
- Increasing QA expectations for health research
- Plenty of protocols
- Industrial processes for data generation

The ACRF International Centre for the Proteome of Human Cancer



# A good (reliable) home for social sciences data:

The screenshot shows the homepage of the Australian Data Archive (ADA). At the top left is the ADA logo, consisting of a stylized green and grey geometric shape followed by the text "ADA AUSTRALIAN DATA ARCHIVE". Below the logo is a dark navigation bar with white text for "HOME", "SUBARCHIVES", "DATA ACCESS", "DATA DEPOSIT", "USER GUIDES", "ADA NEWS", and "ABOUT". To the right of the navigation bar is a search box with the text "Search this site" and a "GO" button. Below the navigation bar, on the left, is a vertical list of sub-archives: "SOCIAL SCIENCE", "HISTORICAL", "INDIGENOUS", "LONGITUDINAL", "QUALITATIVE", "INTERNATIONAL", and "CRIME & JUSTICE", each with a small icon. The main content area features a large green heading "Welcome to the Australian Data Archive" followed by a paragraph: "The Australian Data Archive (ADA) provides a national service for the collection and preservation of digital research data and to make these data available for secondary analysis by academic researchers and other users." Below this is another green heading "New access arrangements for ADA data" with a paragraph: "As of October 2015, ADA data is now available **free of charge to all users**. Please feel free to contact us at [ada@anu.edu.au](mailto:ada@anu.edu.au) if you have any queries regarding this change." Further down is a green heading "About ADA" with a paragraph: "The ADA is comprised of seven sub-archives - Social Science, Historical, Indigenous, Longitudinal, Qualitative, Crime & Justice and International. The archive is managed by the ADA central office based in the [ANU Centre for Social Research and Methods](#) at the Australian National University (ANU)." Below that is a green heading "Accessing Data" with a paragraph: "All visitors to the ADA Data Catalogue can browse and search the catalogue, view study and variable documentation (including frequencies) and download related material (questionnaires, codebooks, etc). Registered users can also analyse and visualise most data online and users who". On the right side of the page is a grey box titled "Recent News" with a RSS icon. It contains three news items: "Enabling free data access through ADA" dated "20 Aug 2014", "A new home for ADA at the Australian National University" dated "20 Aug 2014", and "Release of new Data Management textbook from the UK Data Archive" dated "08 Apr 2014". At the bottom right of the page, there is a "Recent Datasets" section and a page number "16".

YOU ARE HERE: /ADA/HOME

- SOCIAL SCIENCE
- HISTORICAL
- INDIGENOUS
- LONGITUDINAL
- QUALITATIVE
- INTERNATIONAL
- CRIME & JUSTICE

## Welcome to the Australian Data Archive

The Australian Data Archive (ADA) provides a national service for the collection and preservation of digital research data and to make these data available for secondary analysis by academic researchers and other users.

### New access arrangements for ADA data

As of October 2015, ADA data is now available **free of charge to all users**. Please feel free to contact us at [ada@anu.edu.au](mailto:ada@anu.edu.au) if you have any queries regarding this change.

### About ADA

The ADA is comprised of seven sub-archives - Social Science, Historical, Indigenous, Longitudinal, Qualitative, Crime & Justice and International. The archive is managed by the ADA central office based in the [ANU Centre for Social Research and Methods](#) at the Australian National University (ANU).

### Accessing Data

All visitors to the ADA Data Catalogue can browse and search the catalogue, view study and variable documentation (including frequencies) and download related material (questionnaires, codebooks, etc). Registered users can also analyse and visualise most data online and users who

#### Recent News

- [Enabling free data access through ADA](#)  
20 Aug 2014
- [A new home for ADA at the Australian National University](#)  
20 Aug 2014
- [Release of new Data Management textbook from the UK Data Archive](#)  
08 Apr 2014

#### Recent Datasets

16



# European Open Science Cloud

- The European Open Science Cloud (EOSC) aims to accelerate and support the current transition to more effective Open Science and Open Innovation in the Digital Single Market.
- It should enable trusted access to services, systems and the re-use of shared scientific data across disciplinary, social and geographical borders.

Q

LOGIN REGISTER

Twitter LinkedIn

**EOSC**<sub>pilot</sub>  
The European Open Science  
Cloud for Research Pilot Project

HOME ABOUT SCIENCE DEMOS NEWS & PUBLICATIONS EVENTS

DEVELOPING TECHNICAL SOLUTIONS THAT MEET SCIENTIFIC NEEDS

**TECHNICAL CHALLENGES**

ADOPTING NEW, MORE COLLABORATIVE WAYS OF WORKING

**CULTURAL CHALLENGES**

DEPLOYING THE EOSC TO DELIVER OPEN SCIENCE

**SCIENTIFIC CHALLENGES**

# THE RESEARCH DATA ALLIANCE

www.rd-alliance.org

*building the social and technical  
bridges that enable open sharing of  
data*

## 18 FLAGSHIP OUTPUTS

of which 4 ICT  
Technical  
Specifications

## 75 ADOPTION CASES

across multiple  
disciplines,  
organisations &  
countries

## 85 GROUPS WORKING ON GLOBAL DATA INTEROPERABILITY CHALLENGES

*of which 28 WORKING GROUPS  
& 57 INTEREST GROUPS*

## 5,825 INDIVIDUAL MEMBERS FROM 128 COUNTRIES

67% Academia & Research  
14% Public Administration  
11% Enterprise & Industry

## 43 ORGANISATIONAL MEMBERS & 8 AFFILIATE MEMBERS



## Vision

Researchers and innovators openly share data across technologies, disciplines, and countries to address the grand challenges of society.

## Mission

RDA builds the **social and technical bridges** that enable open sharing of data.





# The importance of data for translation

- Research has to be reproducible, so data is needed as ***evidence***
- Research data needs solid foundations so data for research must be ***quality assured***
- Research data is a commodity for ***trust*** building in collaborations
- Research data is an output of research so is best optimized for ***many forms*** of translation

# Data is Transformative

- Governments are not investing in research data to make life easier for researchers
- Investments in research data to enable societal problems to be addressed
- This requires data to be in a form that allows a wide variety of use
- Australia has been investing in this transformation for 10 years now

## FAIR data opportunity/threat

- Publishers increasingly requiring published data
- Funders increasingly recognising research data outputs
- FAIR data protects researchers
- FAIR data builds reputation, partnership, legacy
- The opportunity of FAIR data is to compete on ideas, not on a monopoly of knowledge



## Data Value

- Stronger research – new answers in a rich data environment
- More efficient research – see next
- More trustworthy research - reproducibility
- Stronger partnerships
- More industry engagement – data as a trust builder
- Stronger international engagement on nationally significant problems

# The Value of Open Data

- Data is more valuable if FAIR
- Publications are a reliable means of making information available
- Data has to be reliable
- It has to be provided reliably – through reliable data repositories



# The Value of Open Data Report

The analysis in the report suggests that the value of data in Australia's public research is at least \$1.9 billion per annum and possibly up to \$6 billion per annum – at 2012-13 levels of expenditure and activity.

It is more valuable if it is available through appropriate research data infrastructure

e.g. users of the British Atmospheric Data Centre report an average of 56% of their time working with data – that data is open and with appropriate tools.





## Achieving value: there is a gap

- Research data as a trust builder vs. commercial advantage
- Research data infrastructure is nationally and institutionally based
- System has to recognise value to maintain value – culture change

...to bridge the gap....



# Data Infrastructure Requirements

- Technology – to support scale, complexity, veracity, uncertainty
- Processes – to support the interaction of researchers with research systems, inc. publication, licencing, attribution
- Policy – to determine appropriate outcomes for the expenditure of public money
- People – to ensure that researchers have relevant skills, and data professionals are able to engage

# Research Data Participants

- Researchers: need best data, and simple means to publish and be recognised
- Research institutions: need strong data holdings, and means of preserving data value
- Research data generating facilities: need quality
- National investors: need means of maximising national investment, innovation and jobs
- International: need means of cooperation with all relevant partners

So needs/value are very different in research system



# The Australian Response

- Improved data policy framework
- NCRIS
- Domain initiatives
- Institutional Initiatives
- ARDC
- Emerging skills integrated approach

# NCRIS: Australian Infrastructure Approach

- **Stable** for 10 years
- \$AU150M/annum
- Invests in **collaborative** infrastructure
- Both physical and data
- **Data is infrastructure**
- Separate from research funding
- Substantial national data assets created
- \$20M/annum on data and collaboration services



# Australian Research Data Infrastructure Activity

Nationally:

- Capturing data valuable over long periods in Marine, Astronomy, Earth Sciences, Ecosystems ...for a wide range of research purposes
- Supporting the storage of data
- Supporting the management of data
- Supporting the enhancement of data

Plus lots at research institutions



## Australian Research Data Commons:

- A transformational, sector-wide initiative, working with sector, government, and industry partners to build a coherent national and collaborative research data commons
- To deliver a world-leading data advantage, facilitate innovation, foster collaboration and enhance research translation.

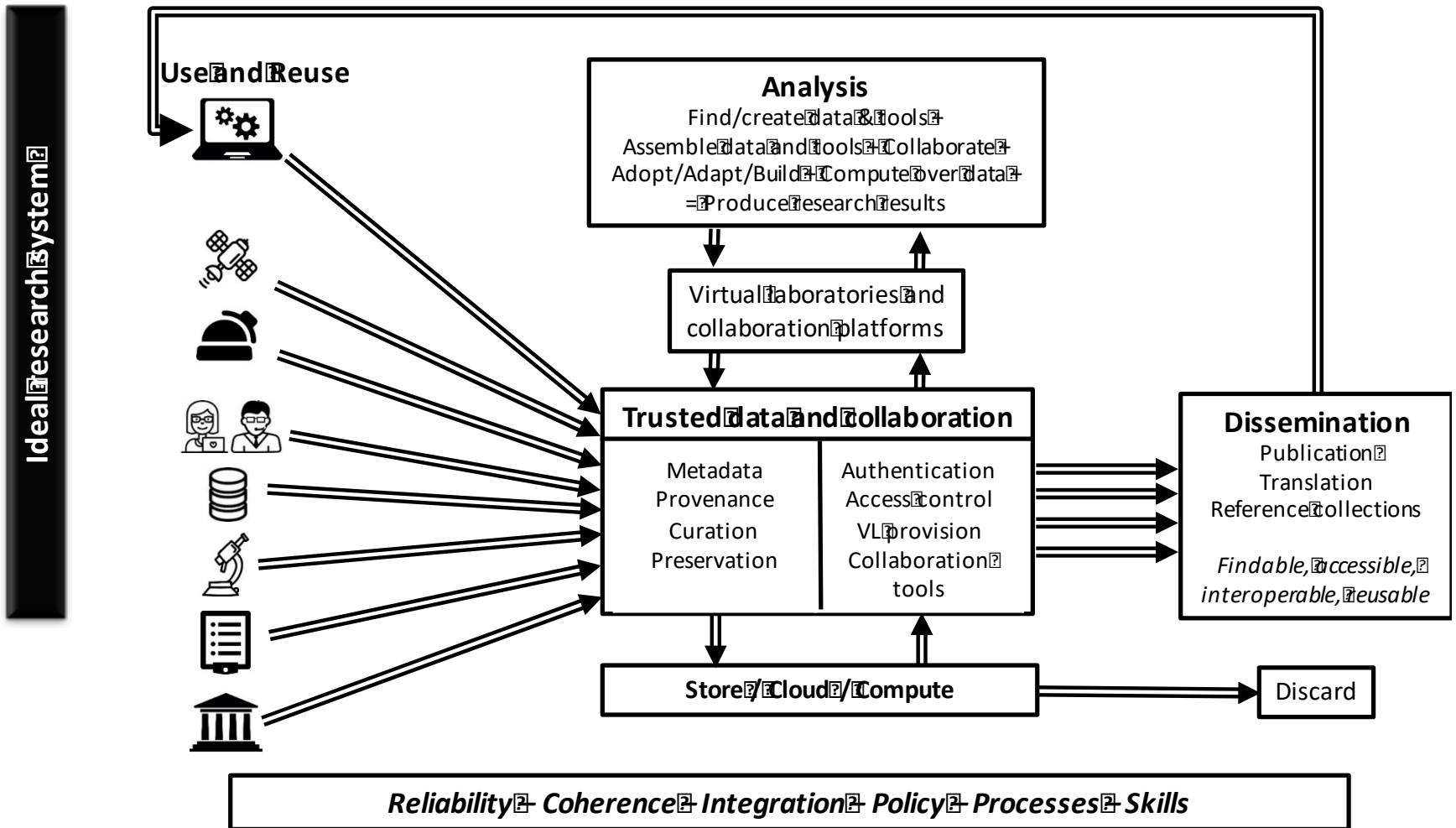


## ARDC activities include:

- Data storage and cloud compute
  - Virtual laboratories
  - Research data management capacity
  - Policy support
  - Community development support
  - Skills support
  - Research Institutional Engagement
  - International Engagement
  - Research data publication services
- ...a systems approach



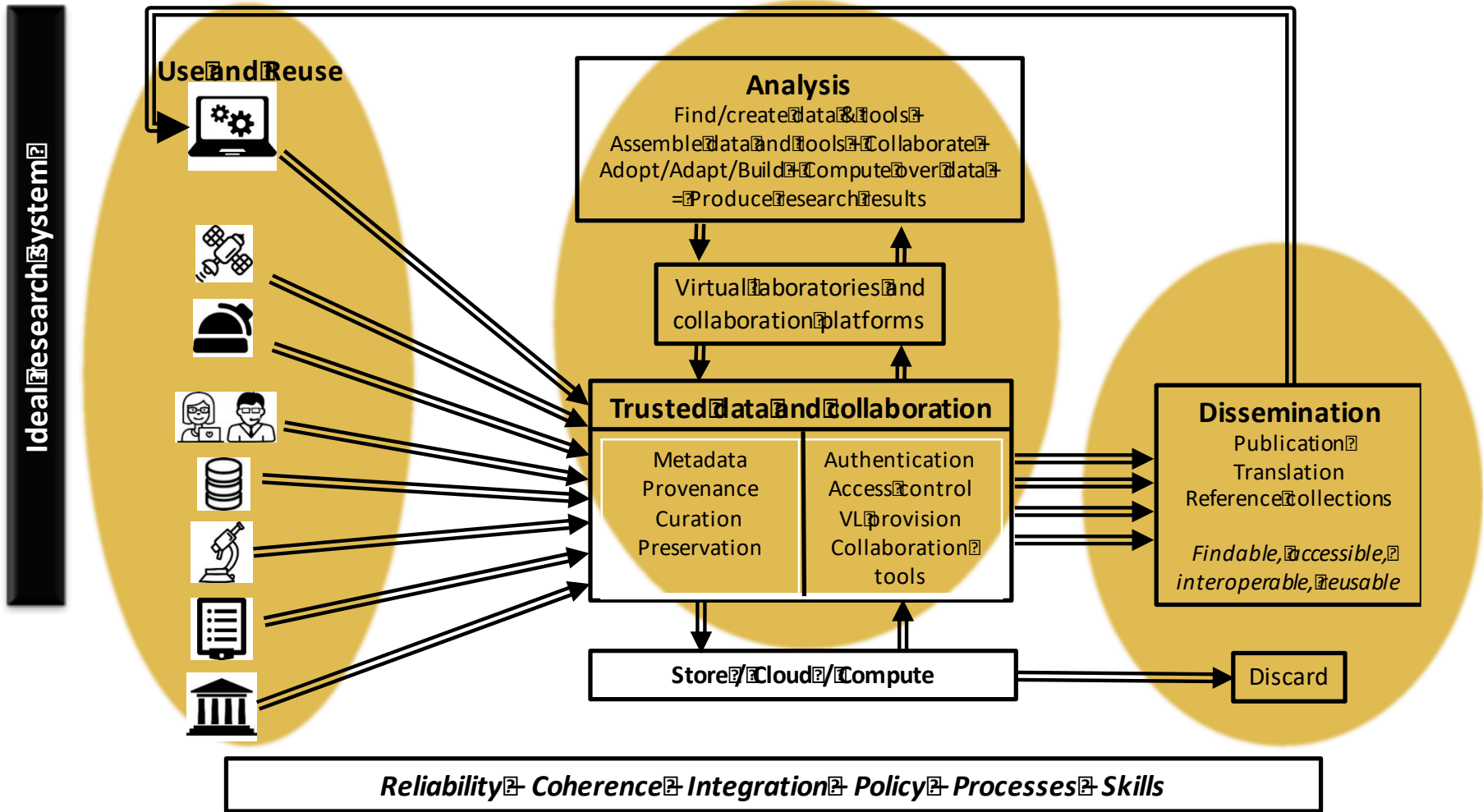
# The research system through a data lens:



Data For Research

Data In Research

Data From Research





## 2013-18 changes

Moved from individual data management/storage/compute initiatives to

- Focused on increasing the value of data to funders, institutions, and to researchers
- Focused on the whole of the research system – an integrated approach to data, data infrastructure, skills, policy, and process
- Strengthened international engagement – mainly through the Research Data Alliance



# Reflections

- Rather than focusing on the data “problem” – focus on the data assets created through research
- The value of data is not static, it is either more or less valuable

# Reflections

- Data is more valuable if it is FAIR, Open and Quality Assured
  - If it is open – it is used more
  - If it is reproducible it can be used as evidence
  - If it is well described – provenance – it can be used for more purposes
  - If it has explicit licences – Use Open licences as much as possible
  - If it is easy to use, because common metadata schema are used
- Make data as open as possible!



# Reflections

Data is not a technology issue only, or mainly:

- Ensure that effort is put into a **professional** data workforce
- Appropriate policy settings requiring and rewarding best data practice
- Good processes are put in place to support policy
- Ensure technology supports easy research data partnerships – nationally and internationally



# Reflections

Research publications are trusted.

Trusted Data is crucial for a world of open science

- Ensure that trusted data repository services are available
- Pay particular attention to provenance
- Build agreements on what constitutes trusted data
- Recognise the value of trusted data in the research system



## In conclusion:

- The value of FAIR research data that is open and quality assured is very high
- The value is different for researchers, disciplines, institutions, nations, and the public
- Use policy, skills, technology together – take a “whole of system” approach
- It is very important to establish international partnerships





Australian Research Data Commons

# Thank you

**NCRIS**  
National Research  
Infrastructure for Australia  
An Australian Government Initiative

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

This work is licensed under a Creative Commons Attribution 3.0 Australia License

