

The Australian National Data Service

Ross Wilkinson

For the CAUL, April 3rd 2009

Role of data

- With more data online, more can be done
- The cost of data acquisition is going down – dramatically
- Possible to get new answers from old data
- Increasing focus on problems across disciplinary boundaries e.g. climate, biosecurity

T

Who cares?

- The Australian Govt - \$24M for ANDS
- The Cutler report – our innovation system
- The institutions - today
- The disciplines??

...and researchers should care..

T

Why should researchers care – and why not?

- The Code requires them to
- The role of data citation
- The changing nature of research

.. but the costs often outweigh the benefits, so ANDS needs to help change the equation

Why might you care about ANDS?

- Research has become more data intensive → data management
- Data is increasingly a research output, rather than a research by-product → data infrastructure
- Excellence in research is correlated with size of effort and with data outputs → data preservation
- Effective response to the Code for the Responsible Conduct of Research is best done collectively

Research Data Intensity

- SKA will generate an exobyte per day
- LHC will generate a petabyte a month
- A current generation gene sequencer can generate a terabyte per day
- Sensors will routinely be deployed to generate enormous and varied data streams for all disciplines
- More data is being captured now that cannot be ever captured again
- More data was created last year than can be stored –

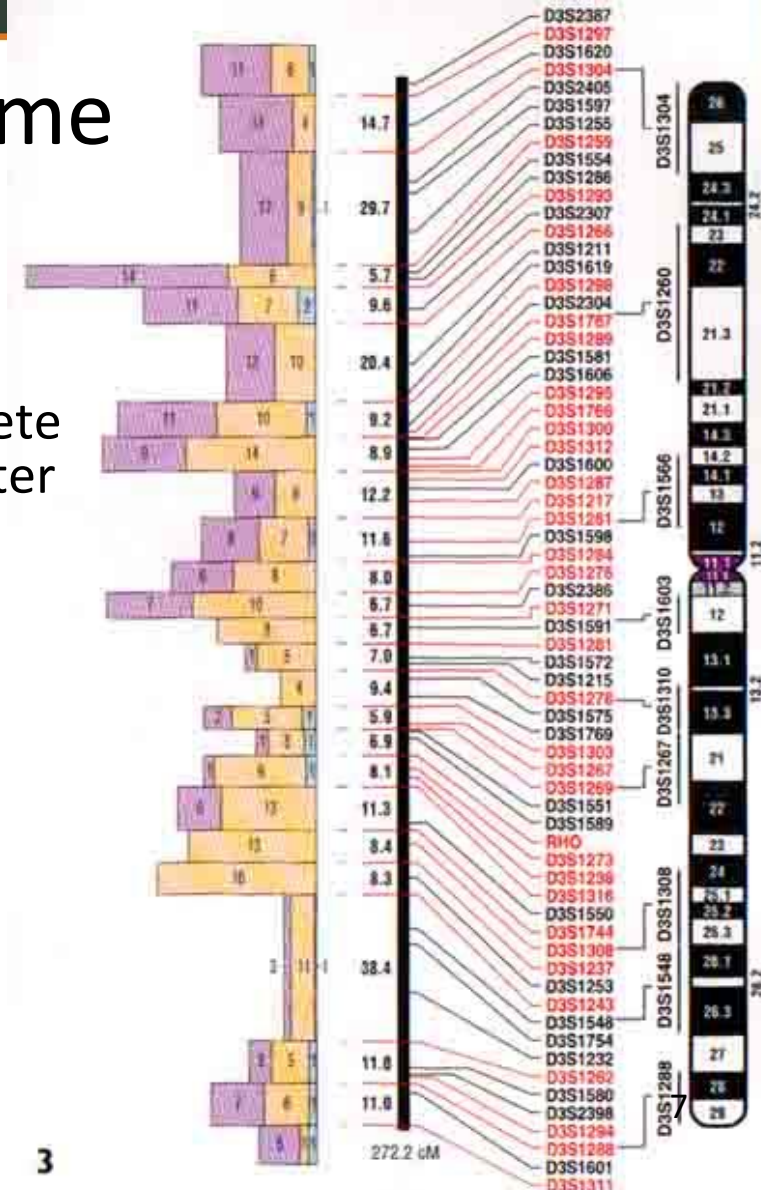
IDC - <http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>

Research Data as Research Product

- The Human Genome project is known for its data
- Currently linguistics is concentrating on data capture rather than analysis as languages expire
- Research data that is used collaboratively can be central to research c.f. TREC
- Hubble telescope data is an output

Mapping the Human Genome

- Took a large team of scientists 10 years to map the 30,000 genes that describe the human body
- In 2007, Craig Venter, published his complete DNA sequence, unveiling the six-billion-letter genome of a single individual for the first time
- The work required a large team using new instruments to produce a large dataset – indeed 2 competing large teams!
- No single lab could have completed this project with available technology in a reasonable time



The Hubble Telescope

- The Hubble telescope launched in 1990
- Increasing focus on cross-disciplinary science
- Observations are proposed, and if accepted, data is collected and made available to the proposers – who then write a research paper
- Each year around 1,000 proposals are reviewed and approximately 200 are selected, for a total of 20,000 individual observations
- The data is stored at the Space Telescope Science Institute
- There are more research papers written by “second use” of the research data, than by the use initially proposed



Excellence

Sharing Detailed Research Data Is Associated with Increased Citation Rate

- 48% of 85 cancer microarray clinical trial publications with publicly available microarray data received 85% of the aggregate citations

Piwowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data Is Associated with Increased Citation Rate. PLoS ONE 2(3): e308. doi:10.1371/journal.pone.0000308

Australian Code for the Responsible Conduct of Research

- It describes the responsibilities of institutions and researchers in the management of research data and primary materials
- Institutions are to retain research data, provide secure data storage, identify ownership, and ensure security and confidentiality of research data
- Researchers are to retain research data and primary materials, manage storage of research data and primary materials, maintain confidentiality of research data and primary materials

http://www.nhmrc.gov.au/publications/synopses/_files/r39.pdf

ANDS is conducting fora on Institutional responses to the data aspects of the code.

To enable the change:

- Shared desire to change
- Professional services – research data analysts, research data carers, professional programmers
- Change partners such as INTERSECT, ARCS, ANDS
- Changed status of research data

ANDS has been structured as four co-ordinated inter-related service delivery programs:

- *Developing Frameworks* – the frameworks that will enable research data producing institutions to capture, manage and share research data;
- *Providing Utilities* –services that reduce the cost of capture and ease the task of discovery;
- *Seeding the Commons* – improving local data capture and populating the data commons; and
- *Building Capabilities* - improving Australia’s capability to manage its research data.

Plus service development for specific disciplines funded through NeAT projects in partnership with ARCS

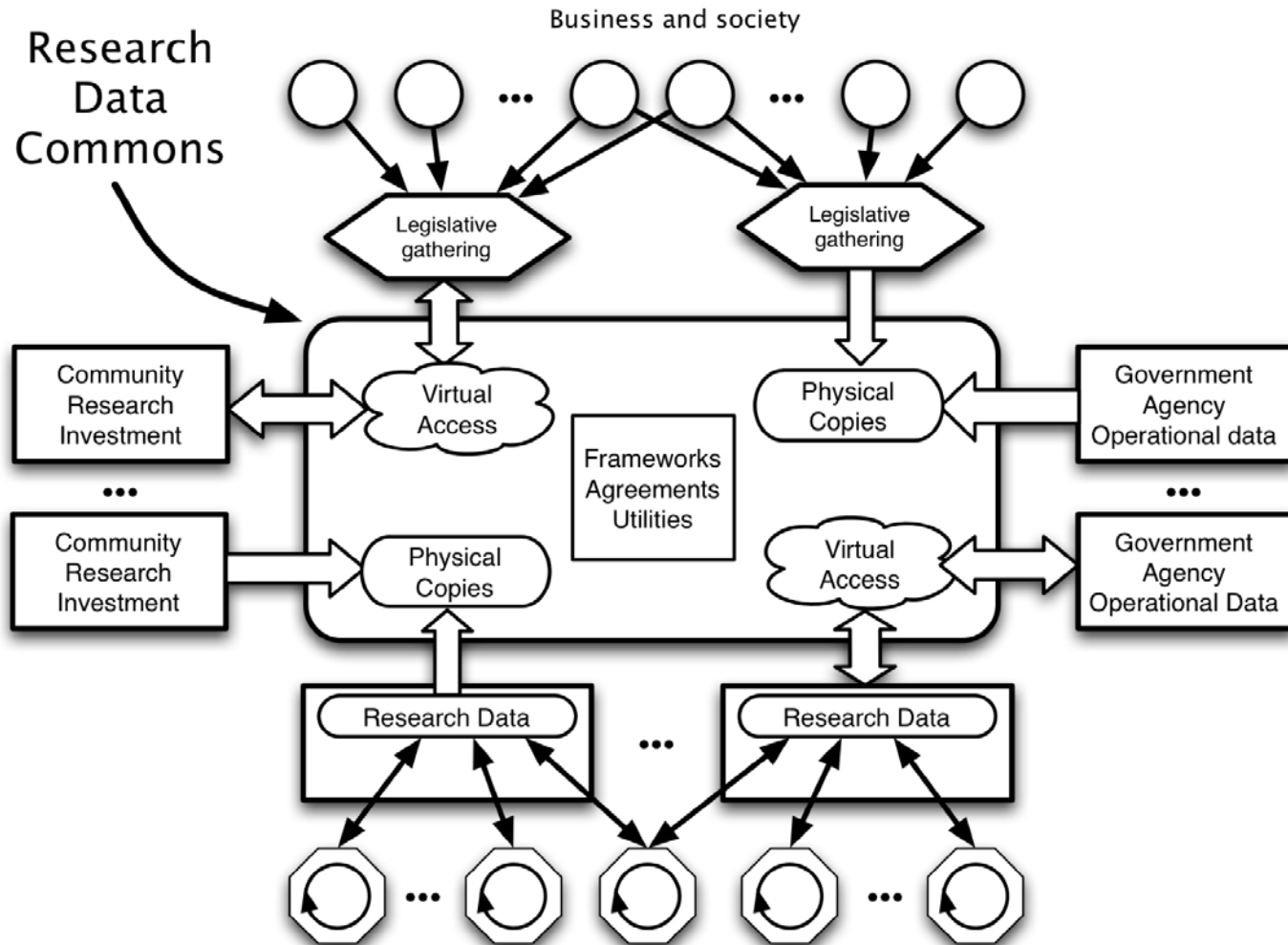
ANDS engagements:

- Our role is to help you achieve your research data ambitions whilst creating the research data commons
- Our mode is to partner at the institutional level and through local partners such as INTERSECT
- We don't have "best practise" to distribute, but sharing good practise
- It works best when the research office, the library and ITS share a desired outcomes

Success:

- An Australian research data commons – a map across Australia's research data, and easy access down into the data
- Skills and tools that enable research data to be managed, shared and cited more easily
- Skills and tools that enable research data to be easily discovered
- Researchers re-using research data more often

The Vision



Collection (Data Commons Page)

Australian Research Data Commons | Collection

http://www.datacommons.org.au/collection/123456.html

ands AUSTRALIAN NATIONAL DATA SERVICE

Australian Research Data Commons

user logged in... LOG OUT

HOME COLLECTION PROJECT FIELD INDEX MY DATA

Data Collection

Title Community-based catch-monitoring of dugongs (Dugong dugon) and green turtles (Chelonia mydas) in the Kaiwalagal region of Torres Strait 2006

Custodian Jillian Grayson, James Cook University

Repository CSIRO Marine and Atmospheric Research Divisional Data Centre

Access MarLIN record : 6843 [DOWNLOAD DATA](#)

Format Database (Microsoft Access), binary and plain text data

Contributor Jillian Grayson, Stephen Ambar, Cyril Stephen, Willie Wigness, Fred Gela, Railey Gibia, Yen Loban, Graham Hirakawa

Description This study trialled community-based catch-monitoring strategies for [dugongs](#) and [green turtles](#) in two communities (Hammond and Thursday Islands) in the [Kaiwalagal](#) region (inner islands) of [Torres Strait](#).
Hunters completed datasheets after each hunting trip and these were collected weekly by Indigenous catch-monitors employed on the project. The catch-monitoring program was evaluated in terms of the reliability of the information collected and the suitability of the method to the communities.

Field [05 Environmental Sciences](#)
[0502 Environmental Science and Management](#)
[050201 Aboriginal and Torres Strait Islander Environmental Knowledge](#)
[050206 Environmental Monitoring](#)

Other Metadata ANDS Collection and Services Registry ID : [876594.org.ands](#)
ANZLIC Identifier : [ANZCW0306006843](#)

SEARCH

Advanced Search

SEE ALSO:

- MarLIN [CSIRO Marine and Atmospheric Research Laboratories Information Network] (Directory)
- Torres Strait Marine Research Repository (Repository)
- DUGONG (ALA)
- GREEN TURTLE (ALA)
- KAIWALAGAL REGION (Google Map)

[Chelonia mydas \(Green Turtle\)](#) (Encyclopedia of Life)

© Commonwealth of Australia | [Privacy Policy](#) | [Copyright](#) | [Disclaimer](#)



LEVEL **1** **2** **3** **4**

Project (Data Commons Page)

Australian Research Data Commons | Research Projects

http://www.datacommons.org.au/project/123456.html

ands AUSTRALIAN NATIONAL DATA SERVICE

Australian Research Data Commons

user logged in... [LOG OUT](#)

[About](#) [Contact](#) [Help](#)

[HOME](#) [COLLECTION](#) **[PROJECT](#)** [FIELD](#) [INDEX](#) [MY DATA](#)

PROJECT

Title Dugong behaviour and the effects of boats and pingers

Data Collection Data Commons ID: [org.ands.876594](#)

Publication Hodgson, Amanda. 2007. "BLIMP-CAM": Aerial Video Observations of Marine Animals. Marine Technology Society Journal 41 (29-43).

Lead Investigator [Amanda Hodgson, CRC Reef Research Centre, University of Sydney and James Cook University](#)

Associate Investigator [Kirsten Dobbs, Great Barrier Reef Marine Park Authority](#)

Program [Conserving World Heritage Values, CRC Reef Research Centre](#)


Project Description This investigation is into the effects of anthropogenic, or human produced, noise on the behaviour of dugongs (species: Dugong dugon).
The objectives of this project are to find out whether boat noise and pingers adversely effect dugong behaviour and could alienate them from their required inshore habitat areas.
The investigation has developed a new method for observing dugongs using the blimp-cam. This consists of a helium-filled balloon (blimp) with a mounted video camera. Controls for the camera and a monitor showing the view from the blimp-cam are on board the boat.
Using this system the investigators have been able to record previously unobtainable information about dugong behaviour including daily behaviour patterns, herd movements, structure and composition, social behaviour, mother and calf relationships and surfacing behaviour.

Field [05 Environmental Sciences](#)
[0502 Environmental Science and Management](#)
[050201 Aboriginal and Torres Strait Islander Environmental Knowledge](#)
[050202 Conservation and Biodiversity](#)
[050206 Environmental Monitoring](#)


SEARCH

[Advanced Search](#)

SEE ALSO:
[BARRIER REEF](#) (Google Maps)
[DUGONG](#) (ALA)



[Data Commons Innovation Map](#)
(related to Amanda Hodgson)

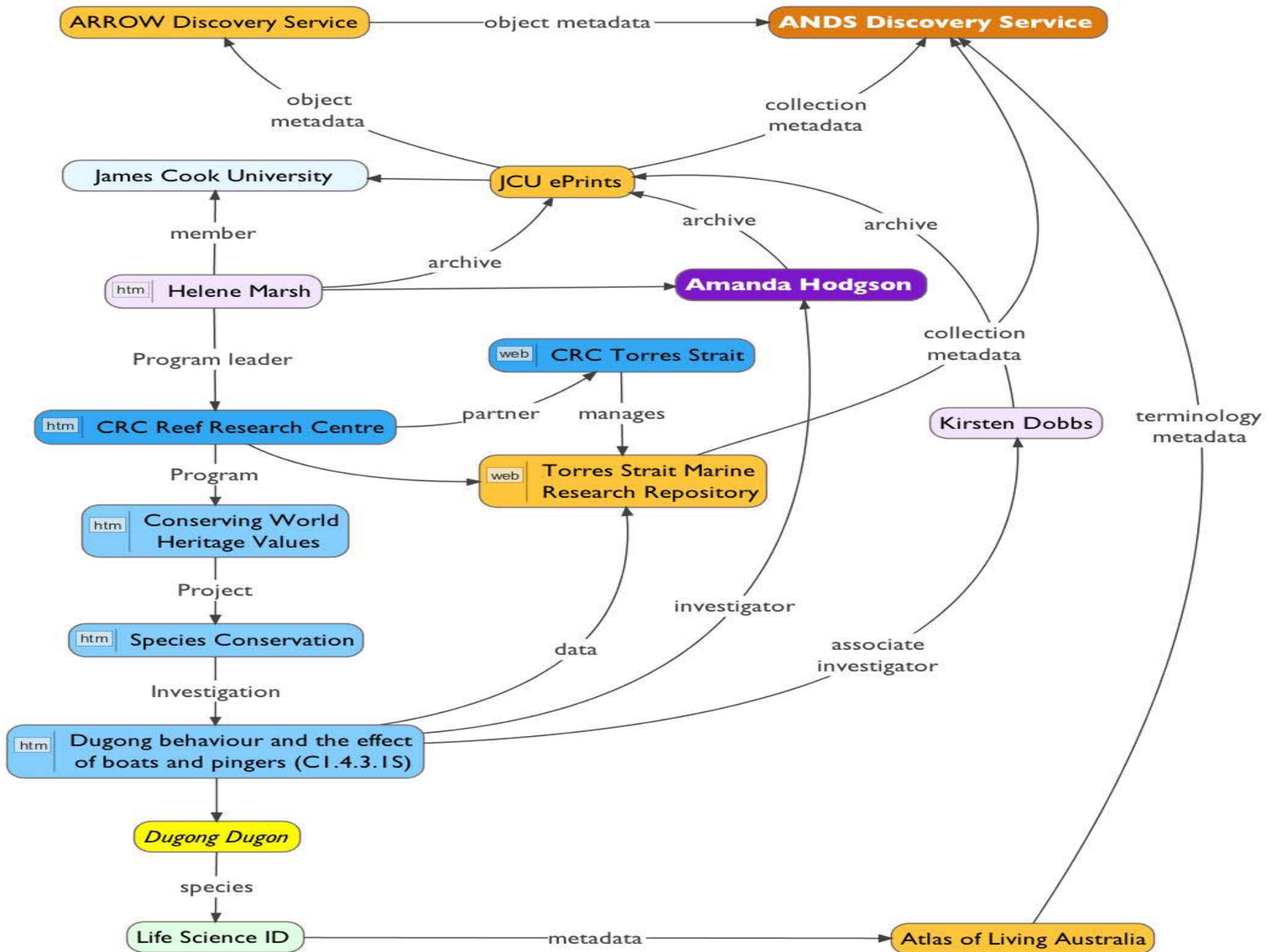


[Getting a Fast Lock on Dugong Locations](#)
(Australian Antarctic Magazine)

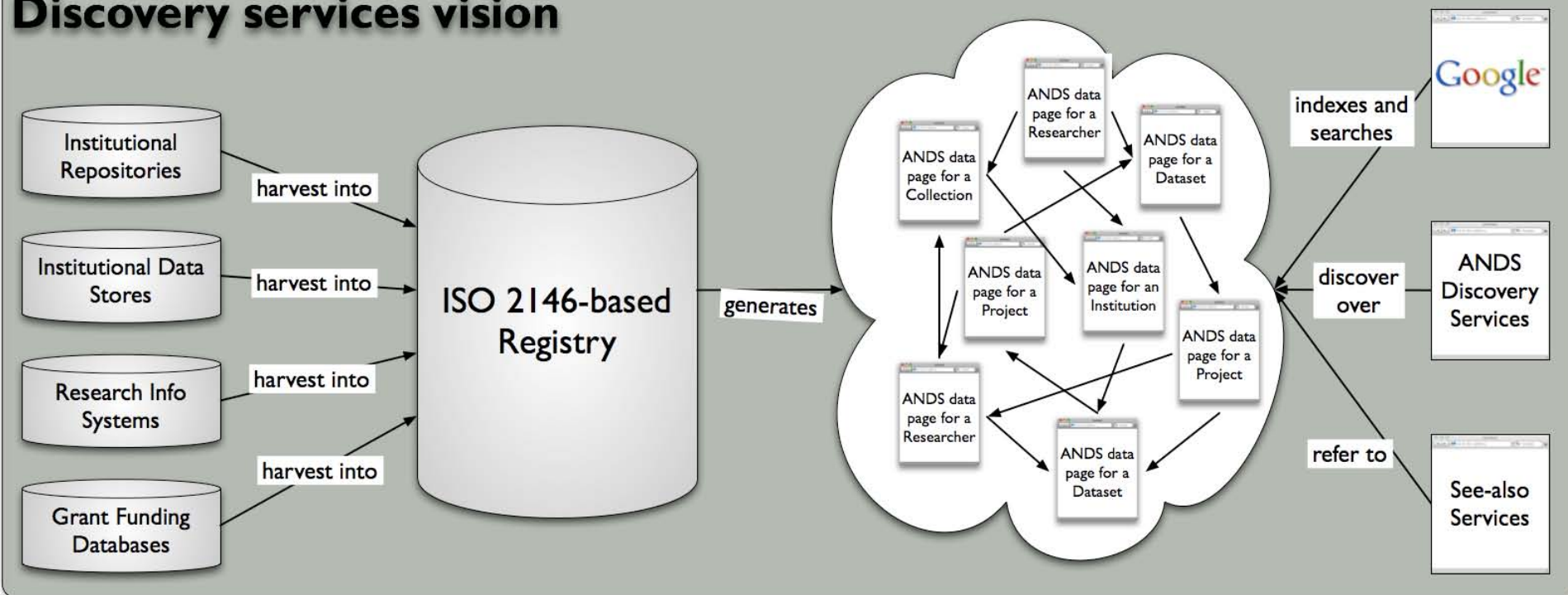
© Commonwealth of Australia | [Privacy Policy](#) | [Copyright](#) | [Disclaimer](#)



LEVEL **1** **2** **3** **4**



Discovery services vision



Early Deliverables

- A national collections registry service
- A national persistent identifier service
- A national discovery service
- Exemplar discoverable data spaces
 - Water?, Crystallography?, Law?, Biosecurity
- Early advise on data capture and management practise
- Partner with Institutional ITS and ARCS for local repositories and storage
- *Not a national storage and management facility*

What do researchers get?

- Locally managed data in a repository
 - Meets required practise, and enables personal re-use
- Persistently identified data
 - Enables data to be a first class research output
- Explorable data
 - Enables data to be explored in the context of the research group, the institute, the collection, the research project
 - A rich discovery environment
- ***ANDS has to help lower the cost of data provision and increase the benefits of sharing***

Success in managing the data driven research world:

More researchers re-using data more often

So we need:

- Lower the costs and raise the benefits
- Data is seen as a first class research output
- Data is seen as an output of first class research
- Partnerships

Thank you.

Frameworks

- Influencing relevant national policies
- Building common understanding of data management issues and solutions across government, research funding agencies, and research intensive organizations
- Encouraging move in favour of culturally-relevant default sharing practices

Utilities

- Building and delivering national technical services to support the data commons
- Examples:
 - Discovery
 - Both “you come to us” and “we come to you” flavours
 - Persistent identifier
 - Collections registry
- Mostly outsourced delivery

Seeding the Commons

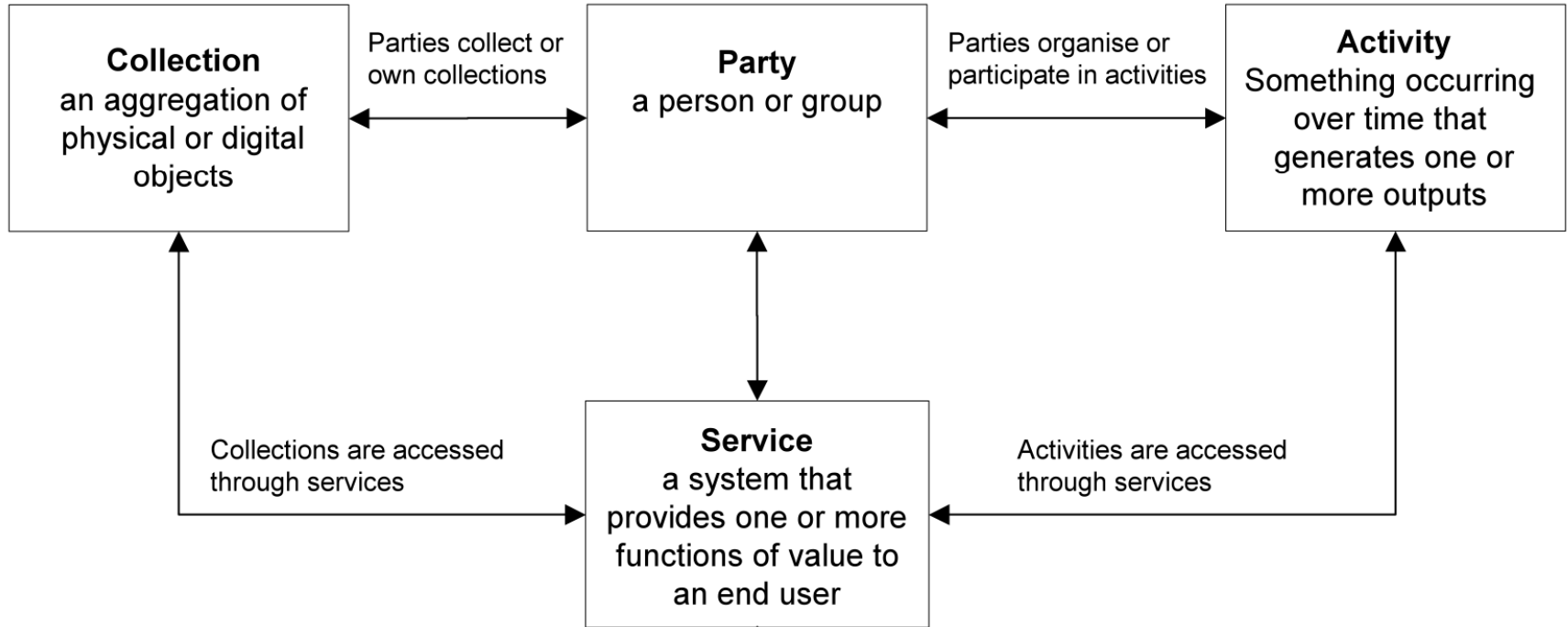
- Improving and standardising institutionally supported repositories
- Supporting Australian research data being routinely deposited into stable and sustainable data and preservation environments
- Working with existing document and data repositories; big focus on content recruitment
- Will need to provide targeted assistance for things ANDS cares about most

Capability Development

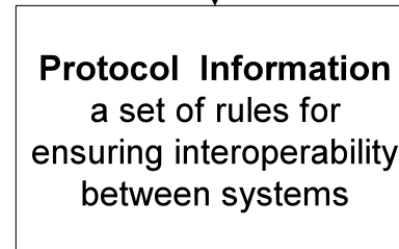
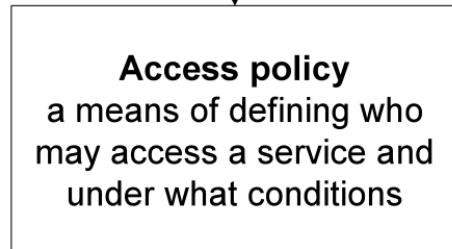
- Assisting researchers to align their data management practices with the needs and outputs of ANDS
- Assisting communities to do things that increase quantity and quality of data available to the data commons
- Learning from existing best practices
- Building culturally-relevant default sharing practices

ISO2146

Registry Object



Services have access policies



Services may be delivered through protocols