

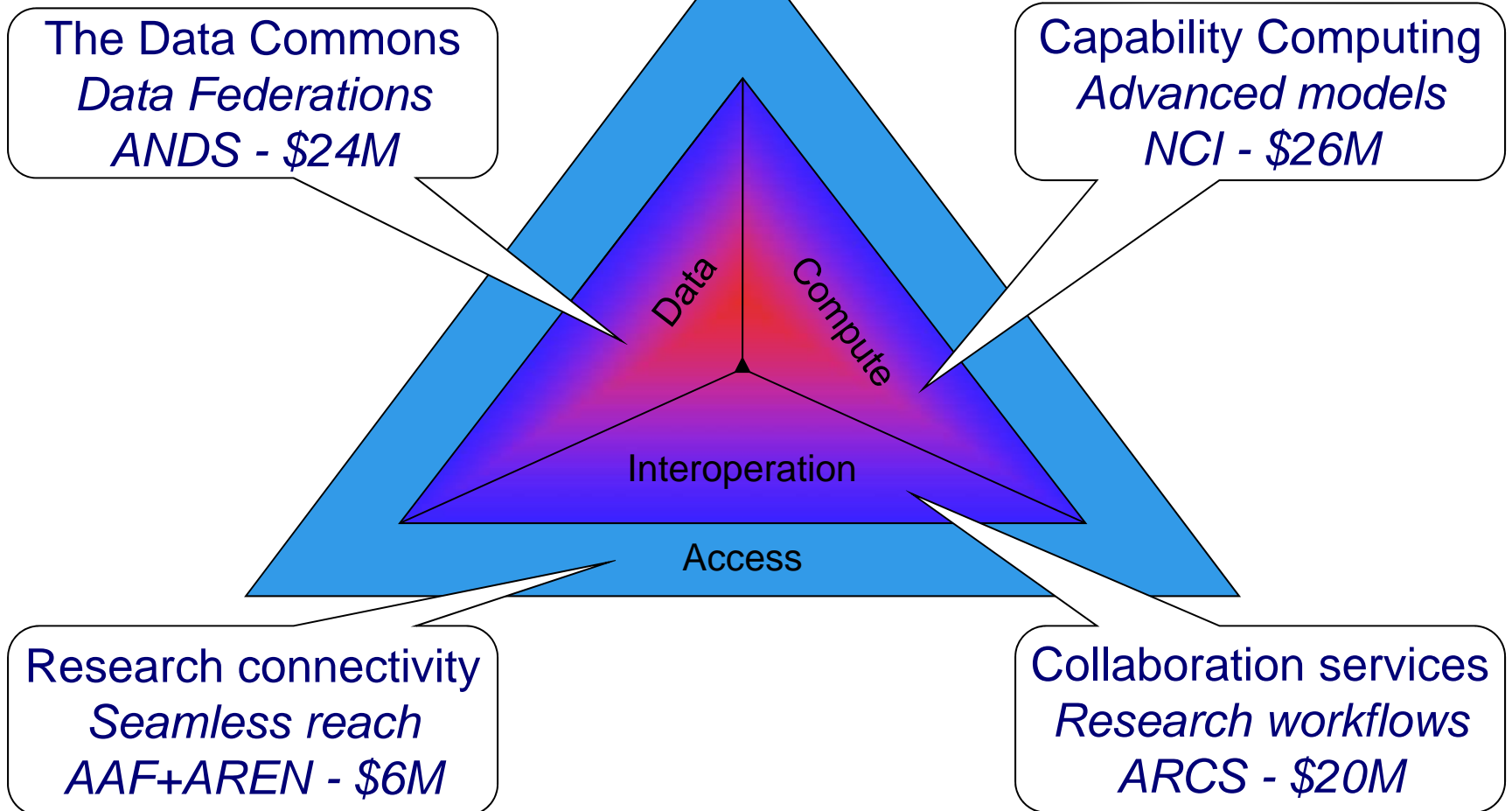
The Australian National Data Service

David Groenewegen
Deputy Director

What is ANDS?

- An initiative of the Australian Government being conducted as part of the National Collaborative Research Infrastructure Strategy (NCRIS) and the Education Investment Fund (EIF)
- A collaboration between Monash University, the Australian National University and CSIRO
- Locations at Monash & ANU

Platforms for Collaboration: Major Investments 2007-2011



The increasing importance 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

Research Data Intensity

- Square Kilometer Array will generate an exabyte per day
- Large Hadron Collider 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 -

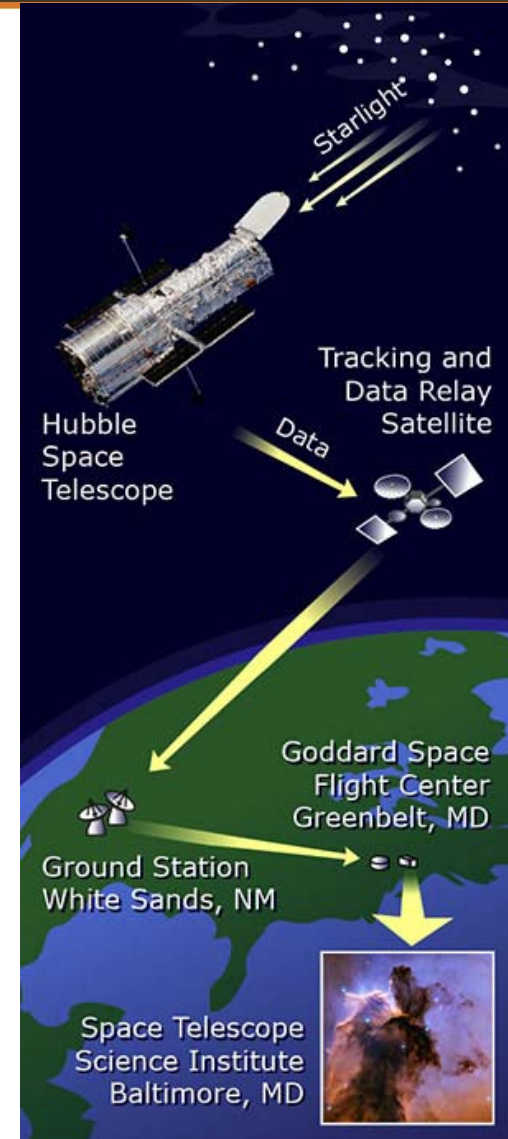
<http://www.emc.com/collateral/analyst-reports/diverse-exploding-digital-universe.pdf>

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
- Hubble telescope data is an output

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

Who cares about data?

- The Australian Government - \$72M for ANDS
- The Cutler report – our innovation system
- The institutions
- The disciplines??

- ...and researchers should care..

Why might an institution 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 AVCC Code for the Responsible Conduct of Research may best be done collectively

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 has conducted fora on Institutional responses to the data aspects of the code - <http://ands.org.au/news/codeforums.html>

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

To enable the change:

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

So how do we plan to do this?

ANDS engagements

- Our role is to help researchers and institutions achieve their research data ambitions whilst creating the research data commons
- Our model is to partner at the institutional level and through local partners
- We don't have "best practice" to distribute, but sharing good practice
- It works best when the research office, the library and ITS work together to meet mutually desired outcomes

Early Deliverables

- A national collections registry service – *Register My Data* (<http://www.ands.org.au/services/register-my-data.html>)
- A national persistent identifier service – *Identify My Data* (<http://www.ands.org.au/services/identify-my-data.html>)
- A national discovery service – *Research Data Australia* – (<http://services.ands.org.au/pages/>)
- *Not a national storage and management facility, not provision of storage hardware*

What do researchers get?

- Locally managed data in a repository
 - Meets required practice, 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***

How will we know if we have succeeded?

- 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**

Thank you for your attention

Questions?

Contact ANDS via:

<http://www.ands.org.au/contact.html>