

Research Data – a Partnership: Researchers, Institutions, National Services, Public Sector Data

Adrian Burton
Director, Services
Australian National Data Service

Australian Govt. Statistical Forum – 30 May 2014

What a data environment is needed for:

- Understanding where and how to build in bushfire prone areas
- Understanding the largest living thing in Australia – the Great Barrier Reef
- The effective use of Australia's soil

??????



eResearch? – Research

- Research needs improved **data**
- Research needs more **computation**
- Research needs better **tools**
- Research needs better **collaboration**
– networks, authentication, tools



Networks

- New optical backbone linking Adelaide, Melbourne, Sydney and Brisbane
- 1-10Gb/s extensions from backbone to Perth and Cairns
- Further leased capacity to Hobart and Darwin
- Regional WANS established in Victoria, South Australia
- Toll free 10Gb/s path to USA and global research networks

- Optical path built to Perth & Geraldton
- 2.5Gb/s - 10Gb/s to Hobart and Cairns
- NBN blackspot to Darwin, regional sites
- Brisbane, Perth, Sydney metro extension
- Victoria, Tasmania, Queensland builds

Tools

Many SII Projects
Investigative and
Developmental

Looming Issues

Data Policy
Data Tools
Data Management
Data Retention

What Software
What Middleware
What AAA solution

- Collaboration Services
- National Data Fabric
 - National Grid & Apps
 - Video Collaboration

- Research Cloud Infrastructure
- Peta scale cloud
 - VM Hosting

Software Research
Tools Development

Virtual Labs & Integration
Software Research Tools

Data

The Australian
National Data
Service

- Data Capture & Description Infrastructure
Data Discovery & Data Use Infrastructure
- Build and populate the Research Data Commons*

Data Storage & Access Infrastructure

Compute

- APAC National Grid
- Apps development
 - Job Launching
 - Server Integration

Pawsey 150 Tf/s

Pawsey
Data
Centre
~ 1500 Tf/s

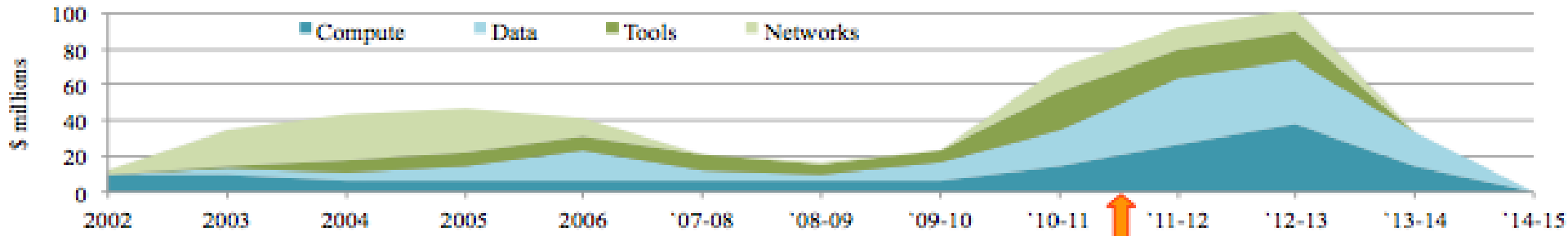
Special Systems 30 Tf/s

NCI
Data
Centre
~ 1000 Tf/s

APAC 1 Tf/s

APAC 14 Tf/s

NCI 150 Tf/s



The need

- Research data intensity is increasing
- Problems are at a larger scale
- A single problem solution is hard to translate
- A research group approach is inefficient and variable
- An institutional approach to data leads to greater research efficiency
- A national approach to data is an enabler for tackling larger problems, more coherency, efficiency and greater synergy
- This requires new partnerships between researchers, institutions, e-research service providers, disciplinary initiatives and national services, the public sector and international initiatives.





Australian Government
National Health and Medical Research Council
Australian Research Council

Universities Australia
www.universitiesaustralia.edu.au

AUSTRALIAN CODE FOR THE RESPONSIBLE CONDUCT OF RESEARCH

Australian Code for the Responsible Conduct of Research

[https://www.nhmrc.gov.au/guidelines/
publications/r39](https://www.nhmrc.gov.au/guidelines/publications/r39)



Excellence

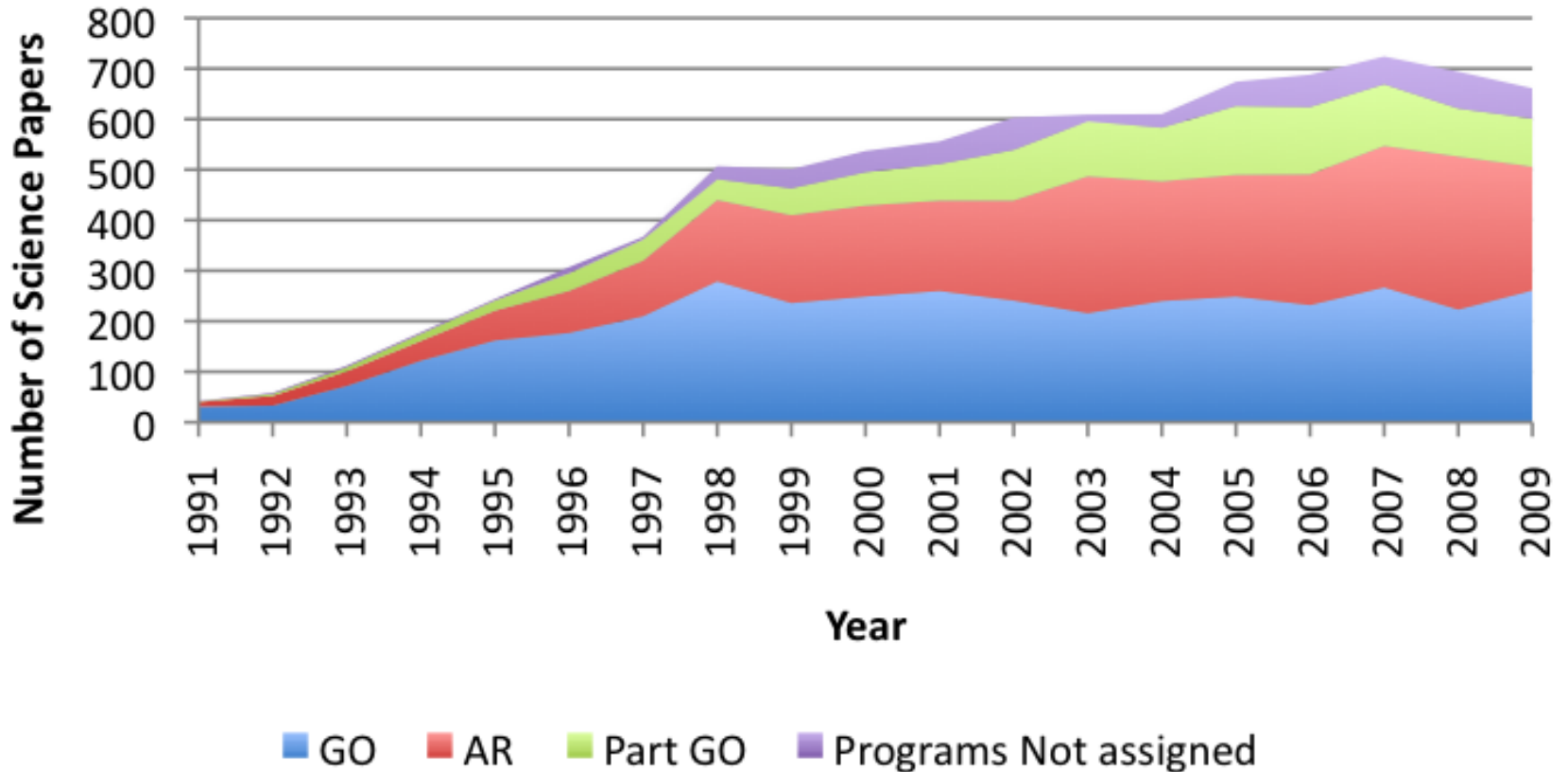
Sharing Detailed Research Data Is Associated with Increased Citation Rate

- 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



HST Publication Statistics



Source: <http://archive.stsci.edu/hst/bibliography/pubstat.html>



Why might a University care about Data?

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



ANDS Infrastructure

- Data Publication
 - National research data portal, registry, metadata harvester, syndication ...
- Identifiers
 - Digital Object Identifiers
 - Handles
- Standard Vocabularies
 - Catalog
 - Machines service
- Support & communities of practice
 - Materials & information sharing
 - Events & Consultancies



Institutional Infrastructure

- Data Capture
- Seed funding
- Metadata Stores
- Data Applications
- Major Open Data Collections



Public Sector Partnerships

- ABS
- GA
- BOM
- Environment
- AIHW
-

