

AUSTRALIAN NATIONAL DATA SERVICE: OVERVIEW AND UPDATE

Andrew Treloar

Director, ANDS Establishment Project

Outline

2

- Context
- Platforms for Collaboration
- Rationale
- Structure
- Establishment Project
- Implications

eResearch Co-ordinating Committee (2006)

Thematic Issues

- Continuing Need for a Focus
 - through national coordination
- Human Capabilities
 - People, skills and understanding
- Linkage of eResearch Resources
 - seamless access to resources
- Access to Data
 - best practice data management and curation
- Structural and Cultural Change
 - evolution of organisational structures and cultures
- Awareness and Support
 - develop researchers' ability to adopt eResearch

Service Clusters

- Data
 - outreach, curation, data management
 - meta-services, location, access, movement
 - practice, providers and users
- Computing
 - capability computing facilities
 - national computing environment
- Interoperation
 - discipline services (tools ((software))
 - user and operations support
 - collaboration services support
- Access
 - the Australian access federation
 - the Australian research and education network



Australian Code for the Responsible Conduct of Research

- The objectives of the Accessibility Framework are also reflected in the new *Australian Code for the Responsible Conduct of Research*
- Published in 2007, the Code replaces the former Statement and Guidelines on Research Practice (1997) from the universities and research funding agencies
- It describes the responsibilities of institutions and researchers in the management of research data and primary materials
- Eg. Institutions are to retain research data, provide secure data storage, identify ownership and ensure security and confidentiality of research data

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



National Collaborative Research Infrastructure Strategy (NCRIS) follows a decade of investment

1997: High Performance Computing Committee

- Established the Australian Partnership for Advanced Computing to provide access to high performance computing capability

2000: Advanced Networks Programme

- Established advanced demonstrator networks

2002: Higher Education Bandwidth Advisory Committee

- Established the Australian Research and Education Network Advisory Committee, and the Australian Research and Education Network

2004: Research Infrastructure Taskforce Report

- Established the National Collaborative Research Infrastructure Strategy Committee to implement a program of strategic investment in research infrastructure

2006: eResearch Coordinating Committee Report

- Outlined an integrated program of skills development and of middleware and computer science research

2007: NCRIS Platforms for Collaboration

- Commitment to an infrastructure program covering computing, data and inter-operation components, and supporting the development of the Australian Access Federation



NCRIS Investments

\$542M** over the five years: 2007-2011

- | | |
|---|--|
| <ul style="list-style-type: none">• Evolving bio-molecular platforms and informatics• Integrated biological systems• Characterisation• Fabrication• Biotechnology products• Optical and radio astronomy• Integrated marine capability• Structure and evolution of the Australian continent | <ul style="list-style-type: none">• Networked biosecurity framework• Population health and clinical data linkage• Terrestrial ecosystem research network |
|---|--|

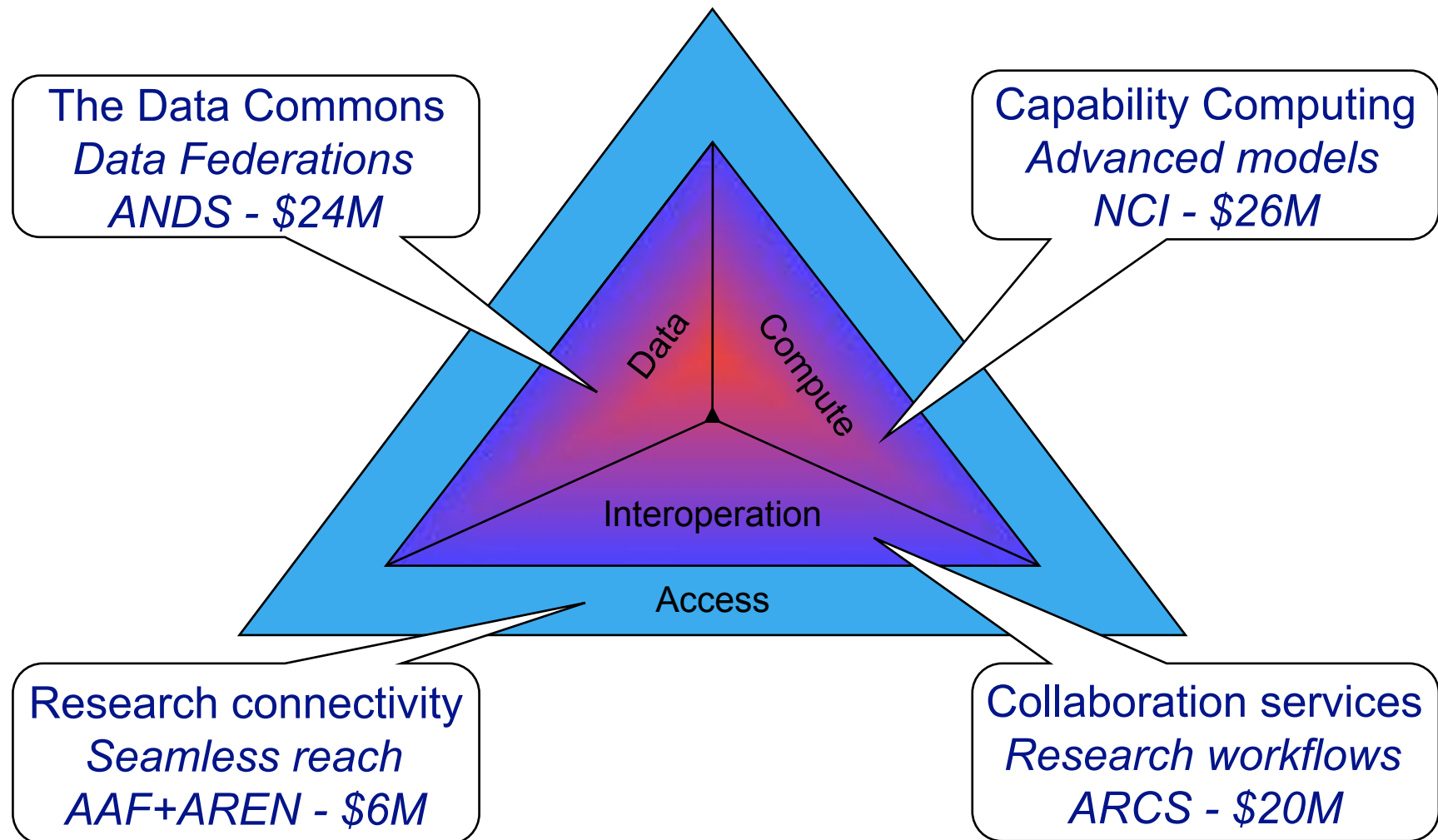
+ *Platforms for Collaboration (allocated \$82 M)*

**Note: to scale to US multiply by 300M/20M = 15



Platforms for Collaboration: Major Investments 2007-2011

7



Towards the Australian Data Commons

8

- Developed during 2007 by ANDS Technical Working Group
- Maps out coherent vision of what needs to be done in the data space
- Available at:

<http://www.pfc.org.au/pub/Main/WebHome/TowardstheAustralianDataCommons.pdf>

Why Data? Why Now?

9

- We are in an era of increasing data-intensive research
- Almost all data is now born digital
- Increasing amount of data generated (semi-)automatically
- *“Consequently, increasing effort and therefore funding will necessarily be diverted to data and data management over time” (Towards the Australian Data Commons (TADC), p. 4)*

Need for standardisation

10

- Software and hardware are getting cheaper, wetware is getting more expensive
- Fixing data management problems is enormously labour intensive and costly
- *“Consequently, standardisation within forms of data and simplification in the frameworks around retention, storage, access and use of data, and the elimination of differences whose resolution requires labour, must be made, if the on-going keeping and reuse of data is to remain affordable” (TADC, p. 5)*

Role of data federations

11

- With more data online, more can be done
- Possible now to answer questions unrelated to reasons why data was collected originally
- Increasing focus on cross-disciplinary science
- *“Consequently greater clarity is needed over control and access to community-funded data, and the means of aggregating, federating and accessing such data are increasingly important” (TADC, p. 5)*

Data unlocks potentials

12

- New scientific instruments
 - ▣ Large Hadron Collider at CERN will generate 1.5 GB of data per second
 - ▣ the Square Kilometre Array (1 EB/day!)
- New scientific Models
 - ▣ The mapping of the Human Genome: A billion DNA letters in a human sequence
 - ▣ Global climate models
- New teams
 - ▣ 195 scientists mapped the genome of the fruit-fly
- New knowledge from unlocked data
 - ▣ Most research from Hubble telescope data was not “first use”
 - ▣ Common data sets unlocked the power of search technology – TREC

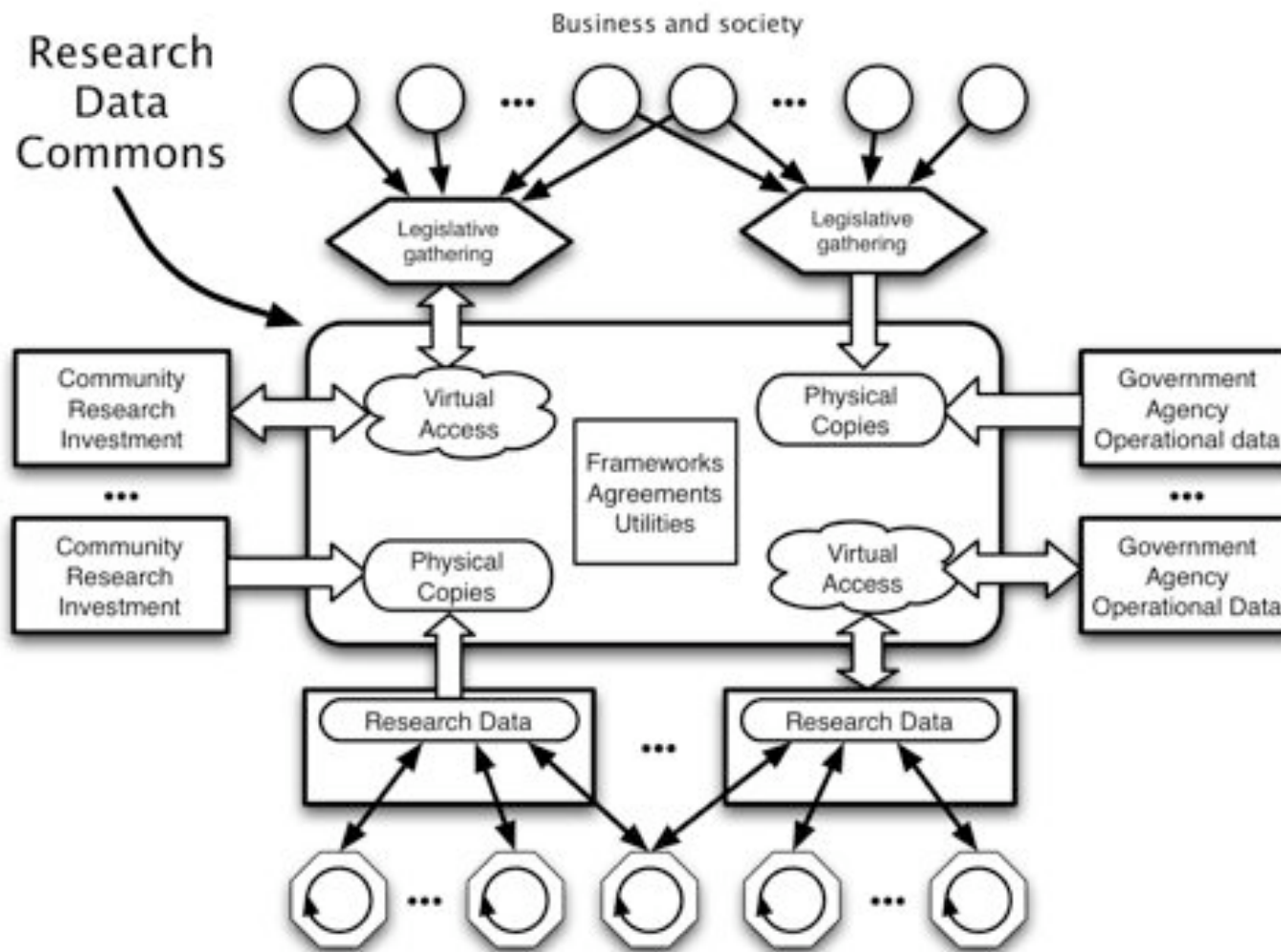
The ANDS Vision

13

- “As a vision, ANDS sets out to transform the disparate collections of research data around Australia into a cohesive corpus of research resources. This transformation would assist the connection of Australian and international data centres, repositories and online collections to enable serendipitous discovery, cross-disciplinary research, and cross-repository workflows.” (TADC, p. 5)

Realising the Vision

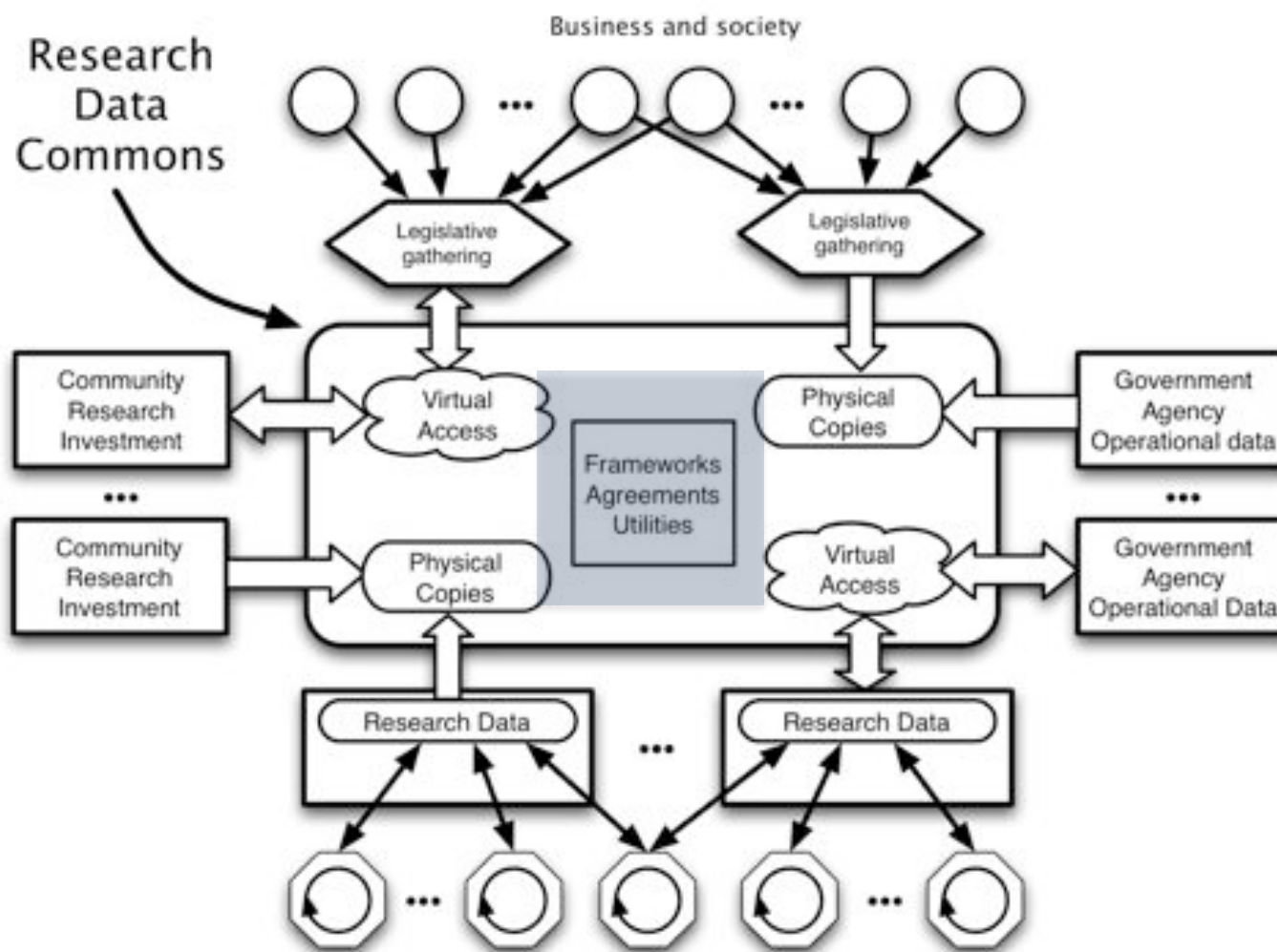
14



Realising the Vision

15

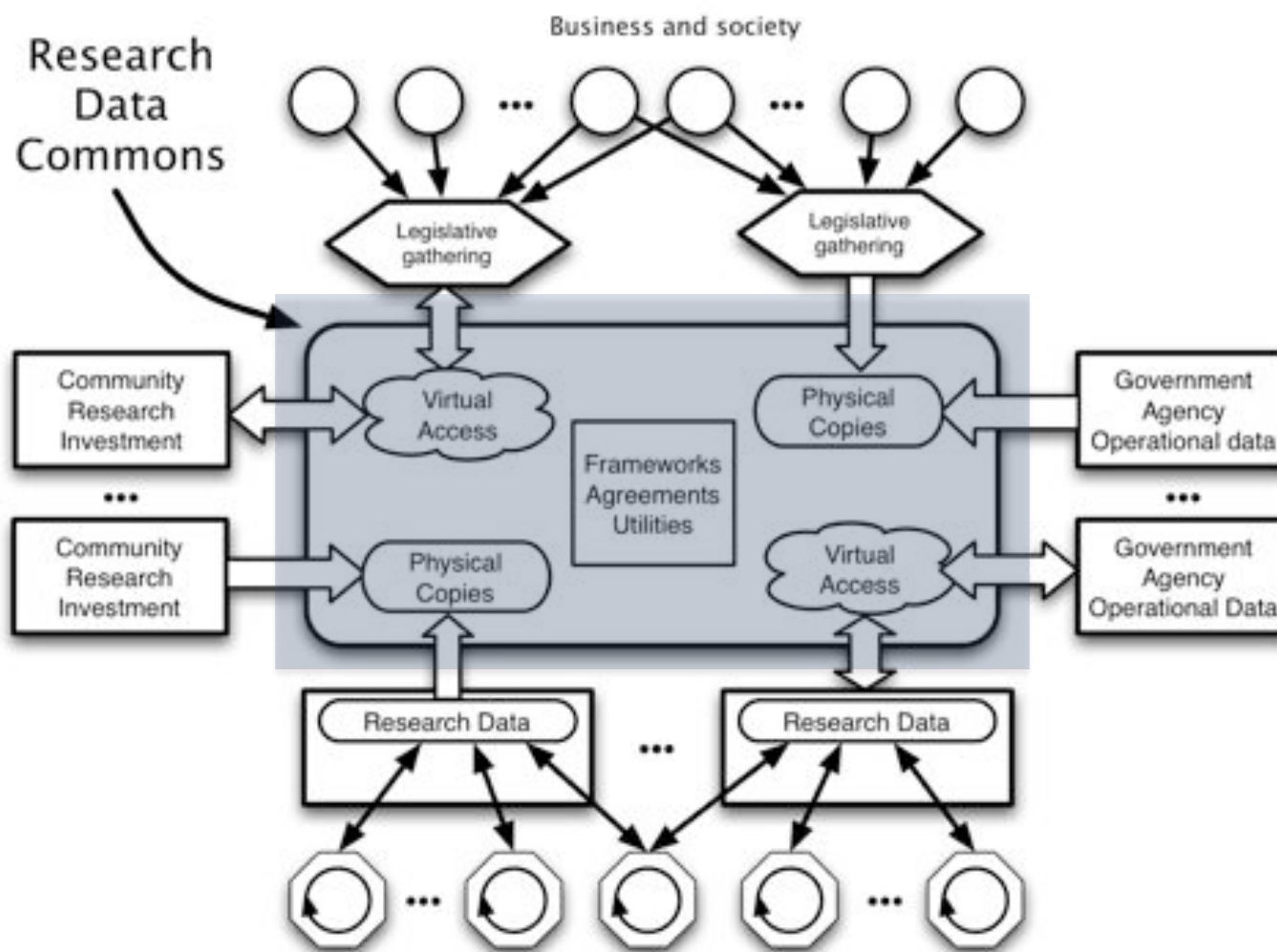
- Develop user and owner frameworks for data commons



Realising the Vision

16

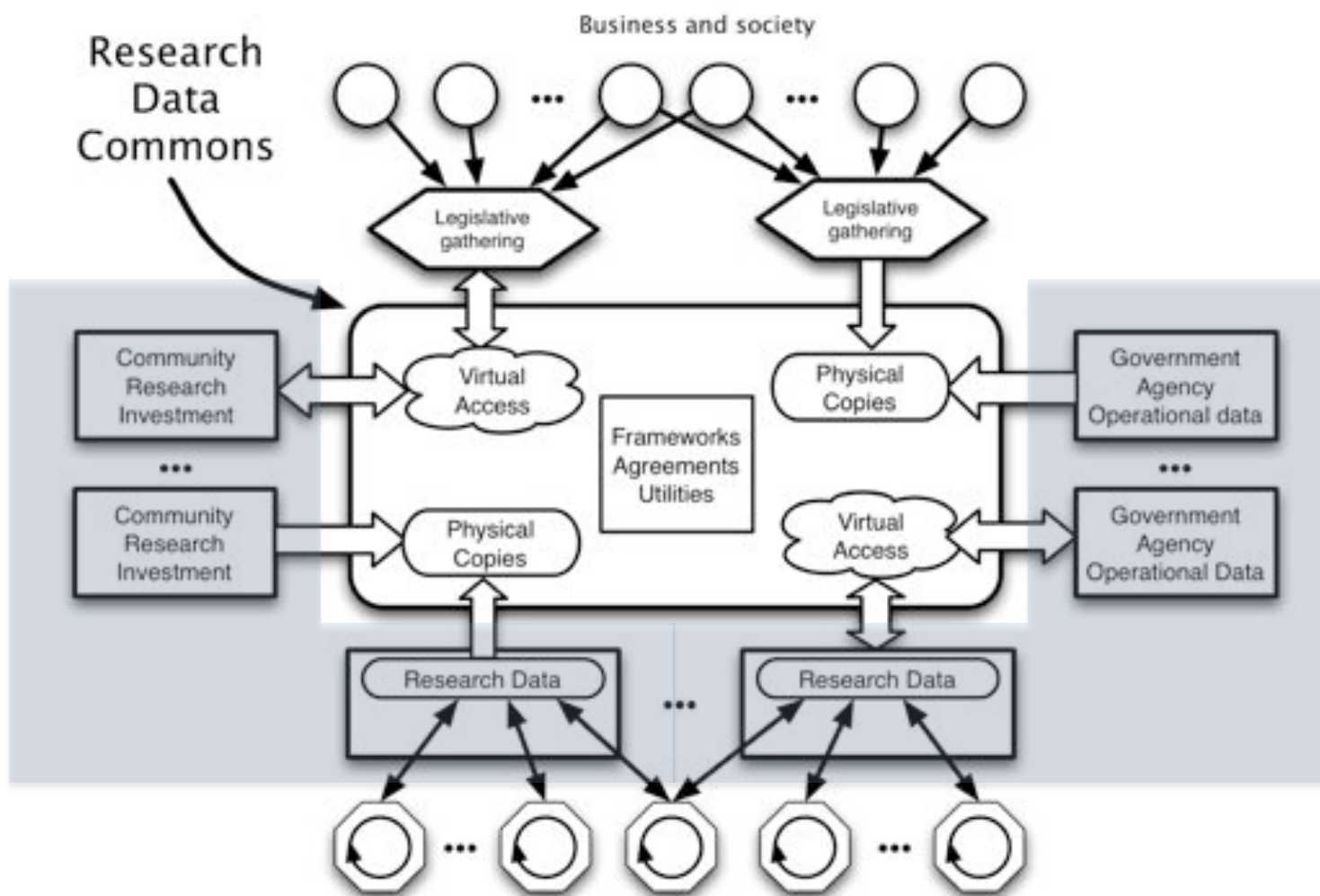
- Develop and operate national registries and discovery



Realising the Vision

17

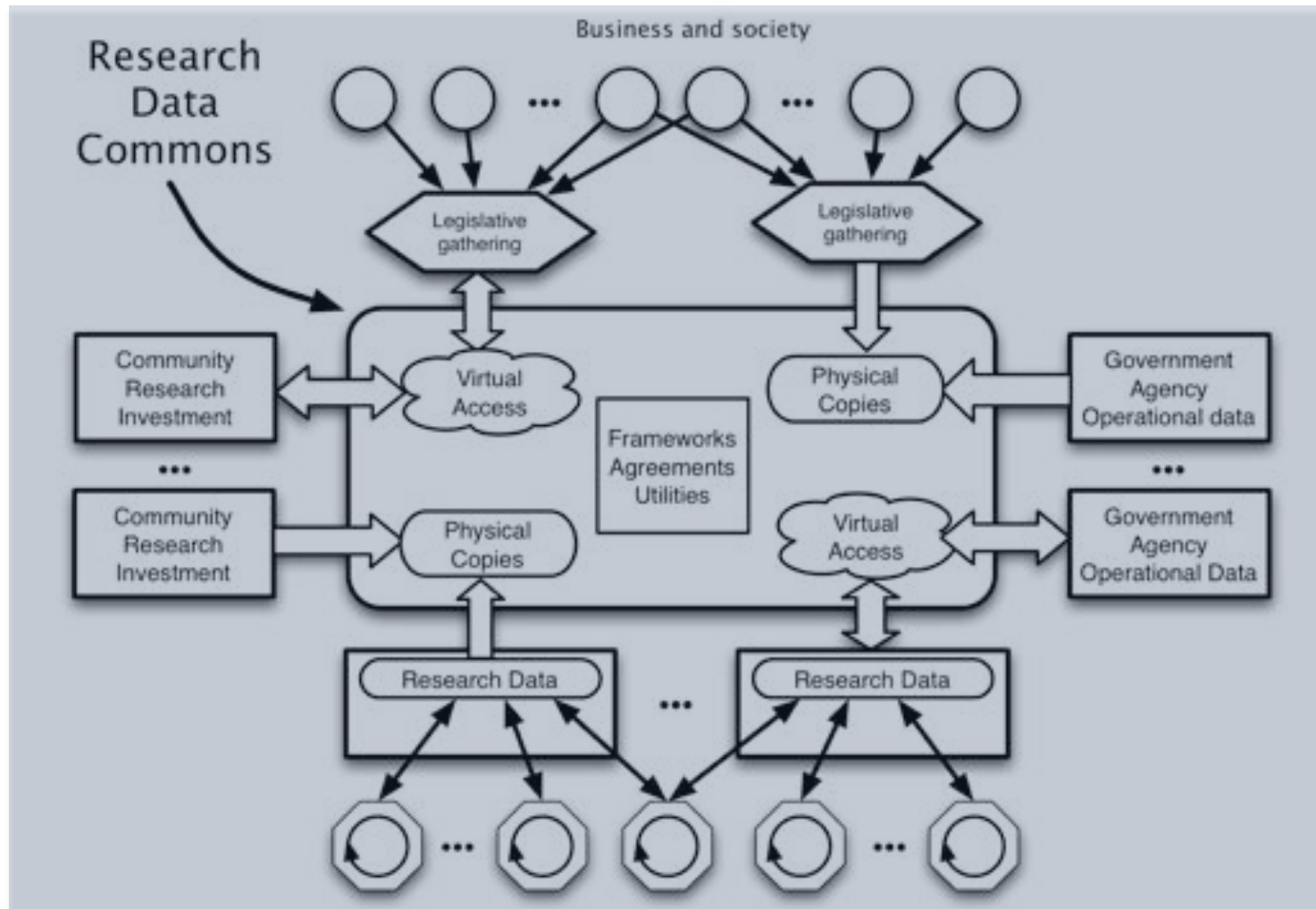
- Seed the commons by connecting existing stores/federations



Realising the Vision

18

- Increase capabilities across sector in data mgt, integration



ANDS Delivery Structure

19

- ANDS has been structured as four inter-related and co-ordinated service delivery programs:
 - ▣ Developing Frameworks
 - ▣ Providing Utilities
 - ▣ Seeding the Commons
 - ▣ Building Capabilities
- Plus development activities funded through National eResearch Architecture Taskforce projects



Developing Frameworks

21

- Influencing relevant national policies
- Building common understanding of data management issues and solutions across government, research funding agencies, and research intensive organizations
- Encouraging moves in favour of discipline-acceptable default sharing practices
- Largely centralised, with some specialised outsourcing



Providing Utilities

23

- Building and delivering national technical services to support the data commons
- Examples:
 - ▣ Discovery
 - Both “you come to us” and “we come to you” flavours
 - Probably a two-step process for some collections
 - ▣ Persistent identifier
 - ▣ Collections registry
- Mostly outsourced delivery
- Some insourced technical framework development



Seeding the Commons

25

- In targeted areas (not enough resource to do everything), working to improve:
 - ▣ fabric for data management
 - ▣ amount of content
 - ▣ state of data capture and management
- Plus, opportunistic content recruitment in first year
- Selection process to identify targets
- Placement of ANDS-funded staff, together with co-investment



Building Capabilities

27

- Improving level of capability for research data management and research access to data
 - Train-the-trainer model
- Providing capability within ANDS for integration of existing systems into Australian Data Commons
- Building community around data management concerns
- Largely distributed

ANDS Establishment Project

28

- Monash (lead), ANU, CSIRO
- Four main deliverables:
 - Formal collaborative agreement to deliver ANDS
 - Funding agreement with DIISR for ANDS
 - Selection process leading to an offer of employment for the ANDS Executive Director
 - ANDS Business Plan for FY 2008/2009
- Originally funded to June 2008; likely extension to September 2008

