Research Methods – The Great Integrator

Rhys Francis
eResearch Futures
rhys_fran cis@icloud.com
Generate new theories using data at scales previously inconceivable

Understand mechanisms impossible to observe or experiment with directly

Undertake novel research studies more extensive than ever before

Do computational modeling, complete data analysis, visualize results

NCI Pawsey

Keep data and observations, describe, collect, share, find, and re-use them

ANDS RDSI

Use new tools, apps, work remotely and collaborate in the cloud

NeCTAR

eResearch Infrastructure

Use new tools, apps, work remotely and collaborate in the cloud

NeCTAR

eResearch Infrastructure

Do computational modeling, complete data analysis, visualize results

NCI Pawsey

Keep data and observations, describe, collect, share, find, and re-use them

ANDS RDSI

Generate new theories using data at scales previously inconceivable

Understand mechanisms impossible to observe or experiment with directly

Undertake novel research studies more extensive than ever before
Computation and modelling
Data is king
Research teams—biosecurity access
# Scale Of National Investment

<table>
<thead>
<tr>
<th>Period</th>
<th>Australian Government Investments</th>
<th>Overall</th>
<th>eResearch</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001 - 2005</td>
<td>MNRF and Strategic Infrastructure Initiative</td>
<td>$400M</td>
<td>$180 M</td>
</tr>
<tr>
<td>2005 - 2011</td>
<td>National Collaborative Research Infrastructure Strategy</td>
<td>$542M</td>
<td>$80 M</td>
</tr>
<tr>
<td>2009 - 2014</td>
<td>Super Science Initiative</td>
<td>$1,100M</td>
<td>$312 M</td>
</tr>
<tr>
<td>2012 - 2016</td>
<td>Continuity (CRIS, NCRIS2013, NCRIS2015)</td>
<td>~$390M</td>
<td>~$75 M</td>
</tr>
<tr>
<td>2016 - 2017</td>
<td>NCRIS 2016</td>
<td>$150M</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>... and beyond</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Building on the national infrastructure platform:

- Many software tools and laboratory integration projects
- Hundreds of data and data flow improvement projects
- Single sign-on for all researchers across the entire infrastructure
- A data commons for the publication and discovery of data
<table>
<thead>
<tr>
<th>Inside the eResearch Infrastructure Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tools</strong></td>
</tr>
<tr>
<td>Provide a dedicated research cloud computing resource</td>
</tr>
<tr>
<td>• associated with data intense centres</td>
</tr>
<tr>
<td>Create a library of re-usable research Apps in the cloud</td>
</tr>
<tr>
<td>Assist research communities create virtual laboratories in the cloud</td>
</tr>
<tr>
<td><strong>Data</strong></td>
</tr>
<tr>
<td>Create a national corpus of published research data assets</td>
</tr>
<tr>
<td>Sustain a sector wide commitment to improved research collections</td>
</tr>
<tr>
<td>Create a network of high capacity, scalable, data storage centers</td>
</tr>
<tr>
<td><strong>Computation</strong></td>
</tr>
<tr>
<td>Construct two buildings for next generation supercomputers</td>
</tr>
<tr>
<td>• fit for purpose over next 10 years</td>
</tr>
<tr>
<td>Install new systems for research use</td>
</tr>
<tr>
<td>• Each providing peta-scale plus computation resources</td>
</tr>
<tr>
<td><strong>Networks</strong></td>
</tr>
<tr>
<td>Expand connectivity to the research network backbone</td>
</tr>
<tr>
<td>• improve regional and metropolitan access</td>
</tr>
<tr>
<td>Extend the fibre backbone - connect east and west coasts</td>
</tr>
<tr>
<td>• connect through to the ASKAP and SKA sites</td>
</tr>
</tbody>
</table>
The next generation to come...

**Research content infrastructure**
- Computational models
- Data Applications
- Researcher Authentication
- Virtual Laboratories
- Collections services supporting and using national data holdings
- Data discovery & publication
- Data access & authorisation

**Shared ICT systems infrastructure**
- Supercomputers
- Shared computing capabilities
- Cloud servers, access methods, tools, middleware and libraries
- High availability servers
- Advanced network connectivity and data movement

**Content Capabilities**
- problem or discipline oriented capabilities
- part of differentiated national and global regimes
- comprising general and specialised data and tools
- supported by multiple institutions
- mapped to appropriate systems infrastructure

**Systems Capabilities**
- provision of national systems that support informatics centric research activities
- coordinated & interoperable
- owned and/or governed by sector participants
- integrated for delivery and support with institutional ICT
Driving the need for new skills...

Long term value resides in
• Data
• Tools and
• Content Skills

ICT systems complexity, especially at scale, can easily dominate e-infrastructures.

Therefore we need to invest more directly in long lived content
Data & Tools & Skills
The Elephants in the Room

New Capabilities & Competencies
- Data/tools migrate over time
- Funding goals need to broaden
- Build or buy content or systems
- Something new is needed

Facilities: Hubble, CERN, EMBL, Synchrotrons, Supercomputers

Research Leadership

Research Curiosity

Research Efficiency

Offices, desktops, laptops, small clusters, data access portals

Research Scale

National & Intl Funding, Supercomputers

Commercial investment, User fees

Unique data, unique software, hand made apparatus, one-off labs

Research grants

Institutional funding
Why these categories endure

Methods

• Different maturity
  => different approaches

Data

• Different natural owners
  ⇒ Research communities
  ⇒ Institutions
  ⇒ Facilities
  ⇒ Preferred supplier

Systems

Networks

• Investment value
  => Inverted to maturity
What have we learned?

2002  Networks  Compute

2007  Networks  Compute  Data  (Tools)
We energised the “value” of data separately

2012  Networks  Systems  Data  Tools
We have to bring compute and data “systems” together
And the maturity of tools is a limiting factor

2020  Networks  Systems  Data  Methods
Future science depends on properly qualified Methods
Methods = Data+Tools on suitable Systems+Networks
The re-use value of data leaves it as a unique asset
Where to from here.....
Obvious questions

• What e-infrastructure does research need?
• How much of what is optimum?
• How do we organise and provide and support its use
• Where consumable or limited in capacity, how do we ration or prioritise its use?
• How do we continue to know the answer to these questions?
More interesting Questions?

• How joined up should all this be? How do we get frictionless infrastructure?
• Do more people care? if so, what do people new to caring about e-infrastructures for research care about?
• Where does the data researchers use come from and where does data they produce go?
• When researchers use data where do they do that?
• How do we share the knowledge encoded in our data and tools with government, health, industry... etc?
• Whose money does what? What does NCRIS money do and not do?
• How can research e-infrastructure benefit from commercial spending?
General outline of the planning activity

- Terms of Reference by end of July
- Project Committee during August
- Invite your inputs on framework topics
- Develop Discussion Paper by October
- Your written responses this year
- Draft framework by March
- Your written commentary by April
- You leverage your international linkages
- You drive the ideas
- You participate in review meetings