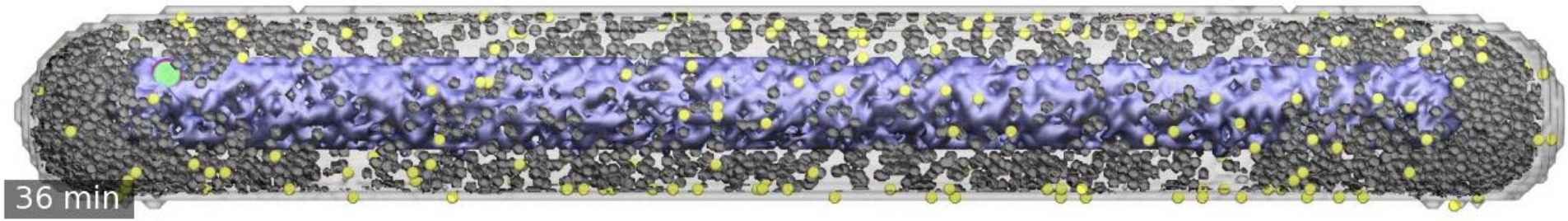
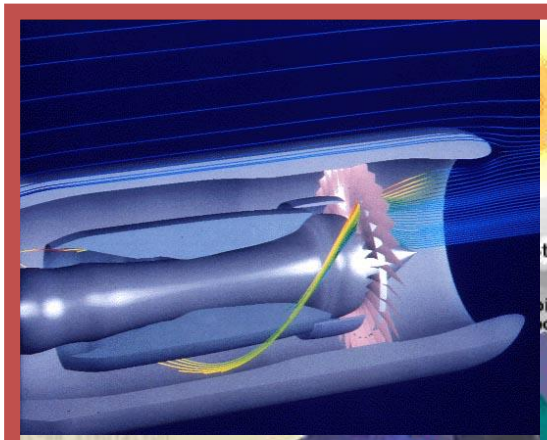
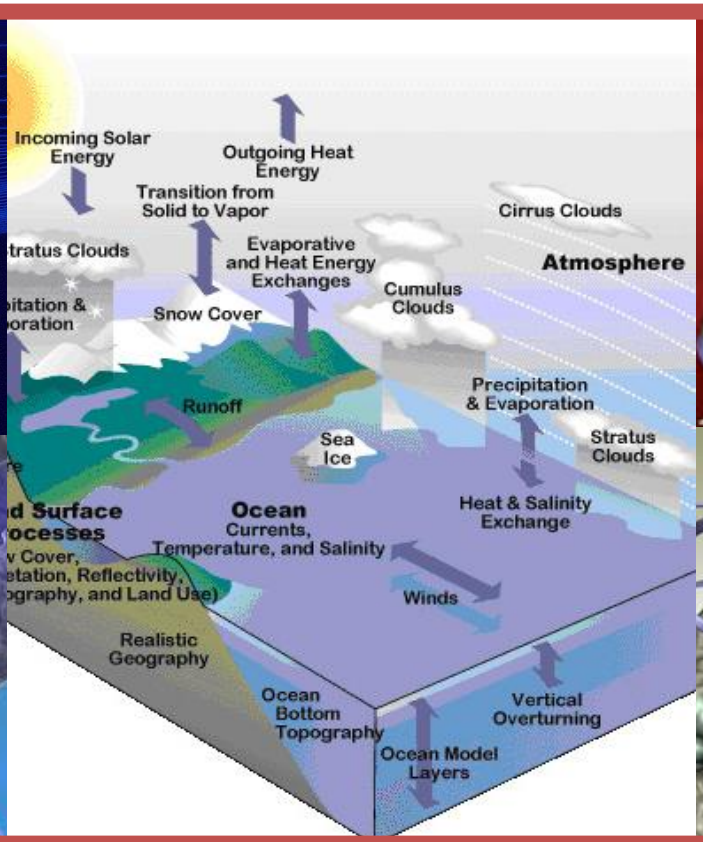
Towards an eResearch Framework

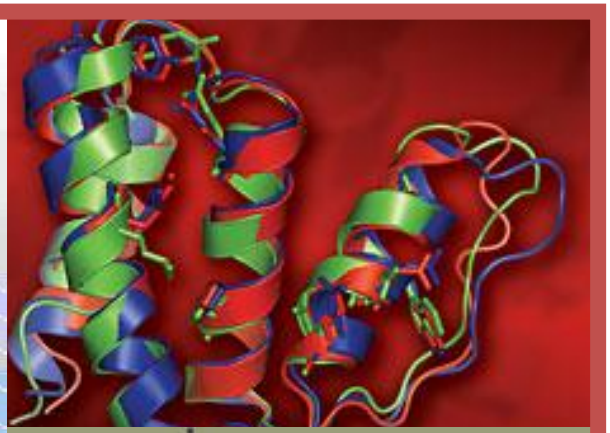
Rhys Francis
rhys_francis@icloud.com

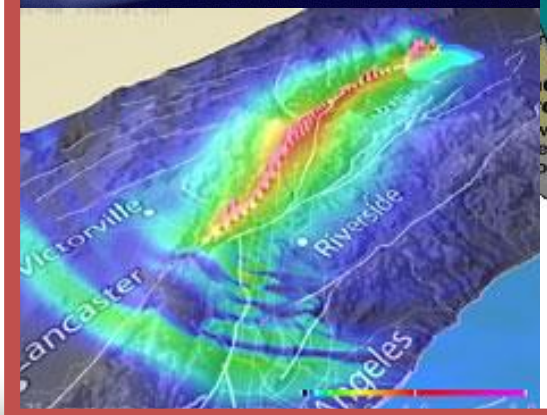
Computation and modelling

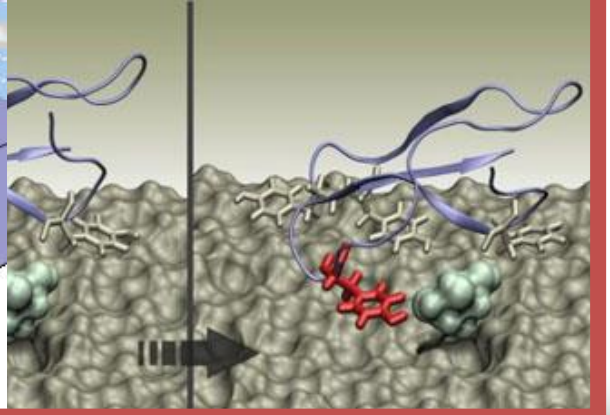




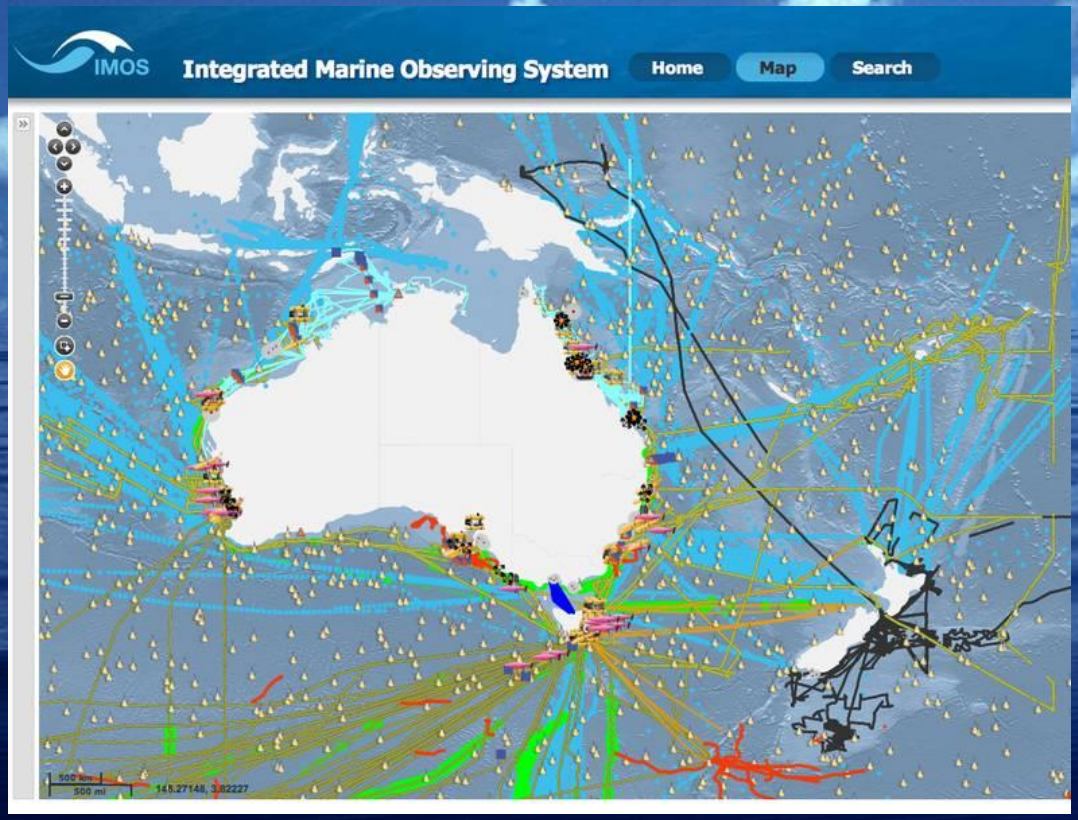


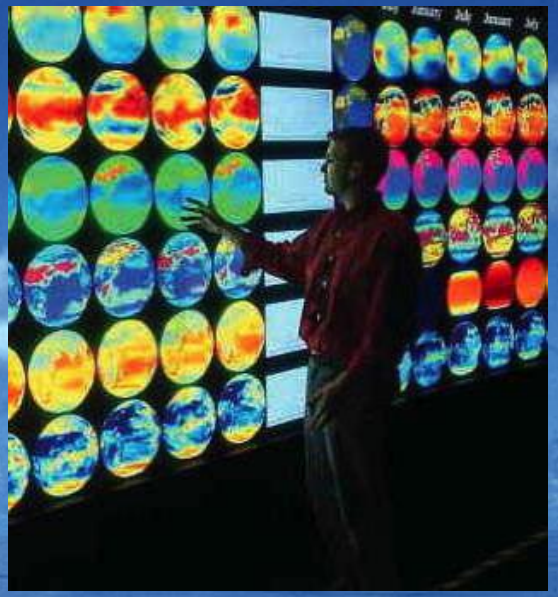






Data is king





Research teams– biosecurity access

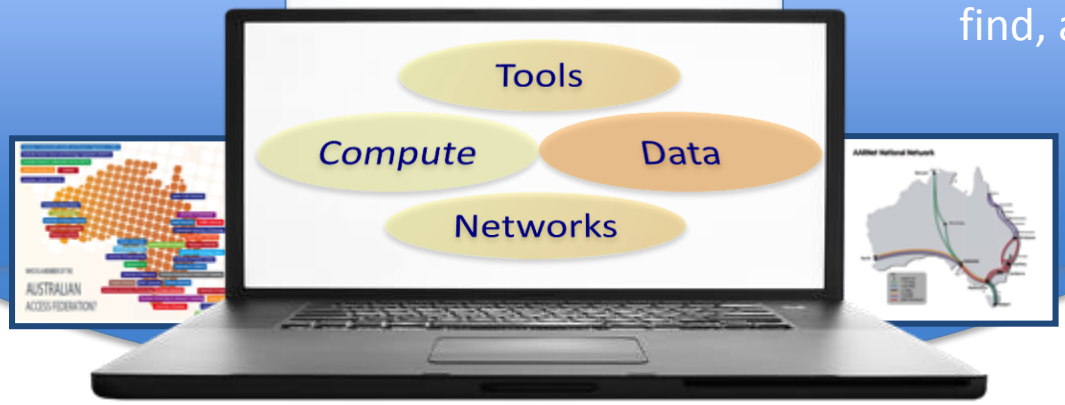
Use new tools, apps, work remotely and collaborate in the cloud

NeCTAR

eResearch Infrastructure

Do computational modeling, complete data analysis, visualize results

NCI
Pawsey



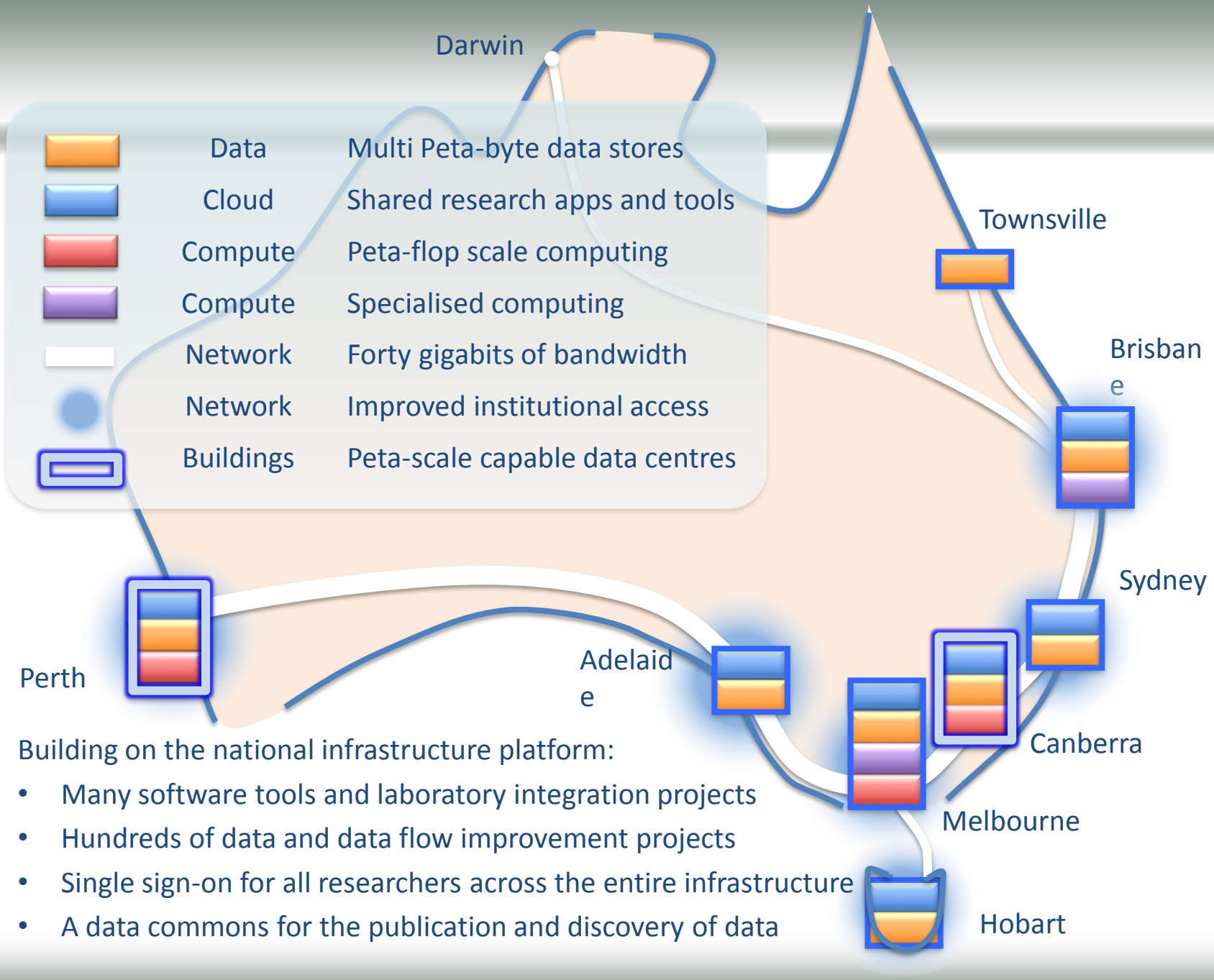
Keep data and observations, describe, collect, share, find, and re-use them

ANDS
RDSI

Understand mechanisms impossible to observe or experiment with directly

Undertake novel research studies more extensive than ever before

Generate new theories using data at scales previously inconceivable



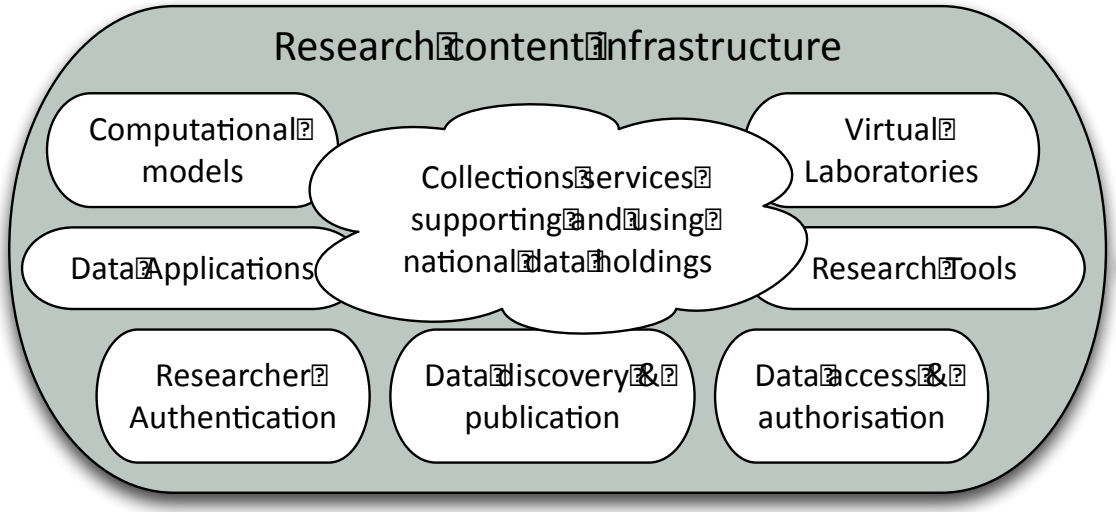
Building on the national infrastructure platform:

- 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

Inside the eResearch Infrastructure Activities

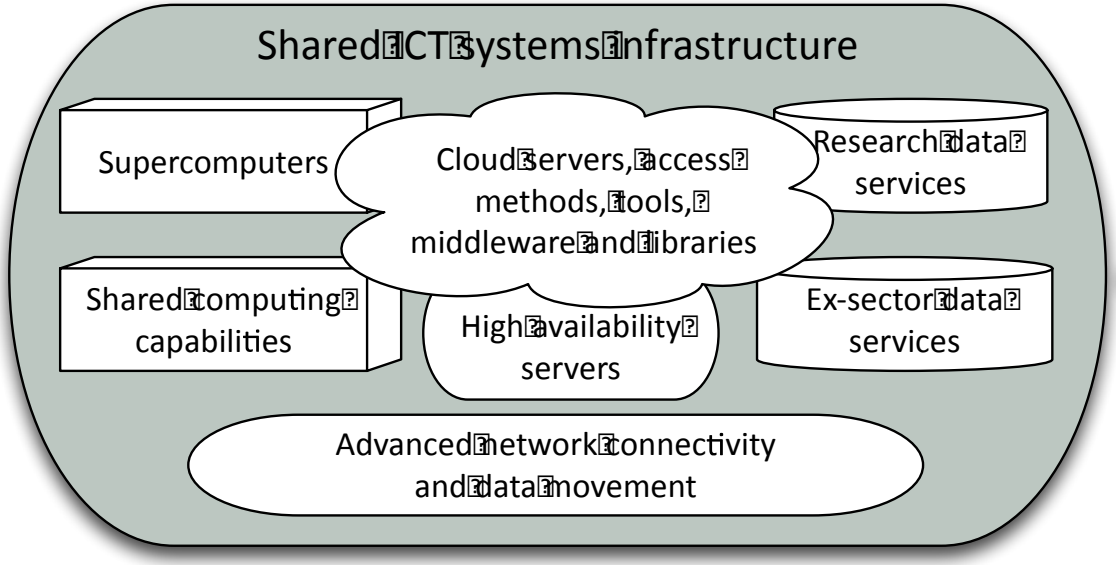
Tools	Provide a dedicated research cloud computing resource • associated with data intense centres Create a library of re-usable research Apps in the cloud Assist research communities create virtual laboratories in the cloud
Data	Create a national corpus of published research data assets Sustain a sector wide commitment to improved research collections Create a network of high capacity, scalable, data storage centers
Computation	Construct two buildings for next generation supercomputers • fit for purpose over next 10 years Install new systems for research use • Each providing peta-scale plus computation resources
Networks	Expand connectivity to the research network backbone • improve regional and metropolitan access Extend the fibre backbone - connect east and west coasts • connect through to the ASKAP and SKA sites

The next generation to come…



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 approrpiate systems infrastructure



Systems Capabilities

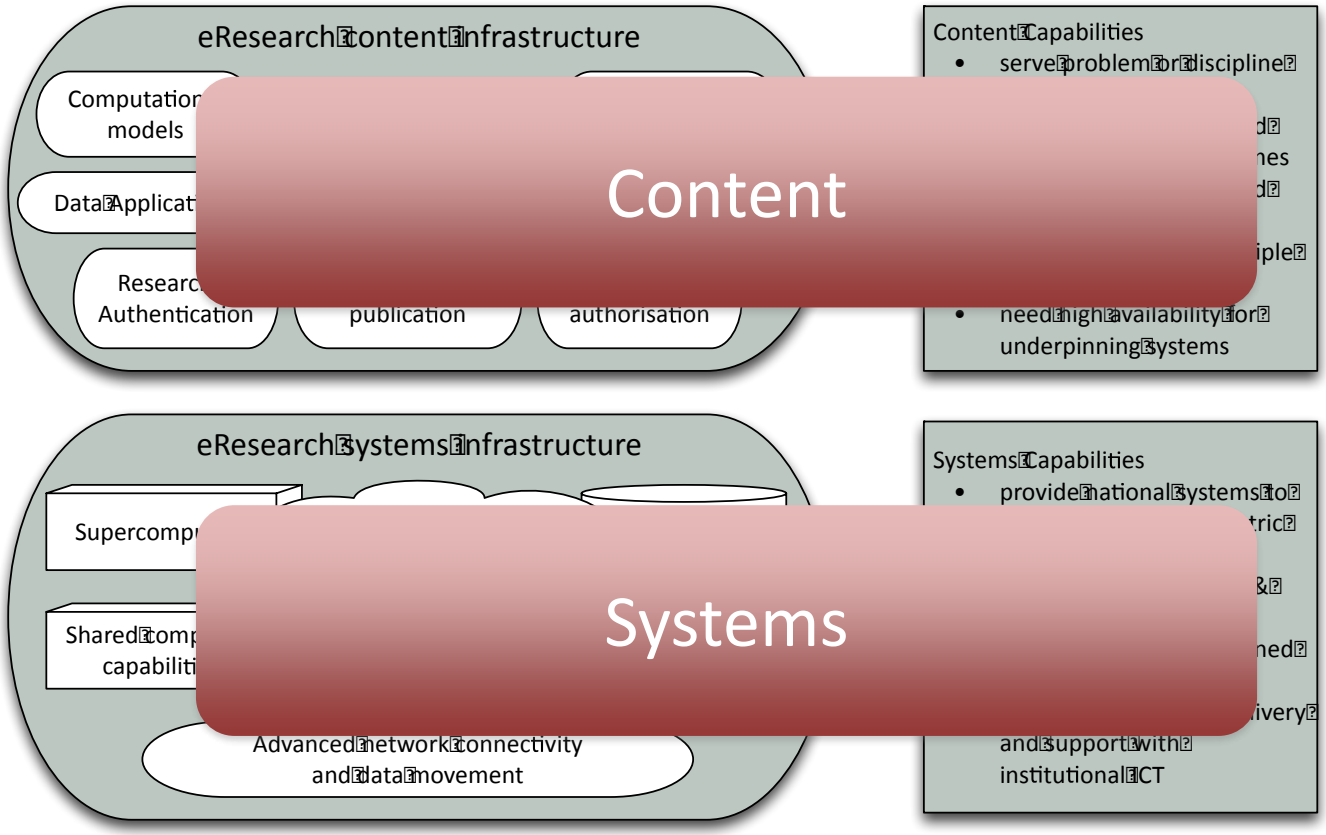
- provision of national systems that support informatics centric research activities
- coordinated & interoperable
- owned and/or governed by sector participants
- integrated for delivery and support with institutional ICT

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

Why these categories endure

Methods

Data

Systems

Networks

- Different maturity
  => different approaches
- Different natural owners
  ⇒Research communities
  ⇒Institutions
  ⇒Facilities
  ⇒Preferred supplier
- IP Investment value
  => Inverted to maturity

What trajectories do we see?

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

The eResearch Framework Project

- Commenced August 2015, Due in April 2016
- Gathered input and feedback – WP First Draft
  - Critical Uncertainty paper generated
  - Key issue – what is the focus of NCRIS eResearch?
- Themes and Issues – WP Second Draft
  - Identified missing themes and issues
- Towards framework– top level – WP Third Draft
  - Identified further missing issues
  - Identified critical ‘head office’ direction

What have we learned

- The eResearch capabilities are valuable and in demand.
- Network and compute requirements continue unabated.
- The impact of digital data on research is a third pillar of value.
- Data must be connected with methods and publications.
- Mechanisms to develop skilled support staff are necessary.
- Researchers will need to include new skills in their teams.
- A build up of skills at the point of service delivery and in the service itself is required.
- Changes to culture and practice are themselves important challenges.
- eResearch and digital data infrastructures are part of improving national research competitiveness.

Feedback

- Conceive of the desirable future and engineer a pathway to get there.
- Establish a high quality guidance process others can observe and engage with.
- Establish a mechanism for effective accountability.
- Set out to make concerted strategic contributions to other parties.
- Use structures/processes appropriate for an operating infrastructure.
- Develop and apply more uniform research impact, value and quality measures.
- Address the expertise and skills issues in a more visible and direct way.
- Promote the ‘right things’, make those ‘right things’ easy, and deliver those things.

The big question

- The question has never really been about how to improve each eResearch project
- The question has been about how to improve the eResearch idea
- We have gone a long way round to what we actually knew at the start

There are two different ideas

There are two primary impacts….

eResearch-Green:

- is about the performance of high impact and nationally prioritised research;
- provides for the competitive use of scarce resources specifically building on high performance computation modelling and data reduction and analysis; and
- is a computational simulation and data generating and data using infrastructure.

eResearch-Gold:

- is about the achievement of a borderless accessibility of quality research data;
- supports the co-operative possession of digital data assets in a way that allows for the maximum possible use and re-use of them; and
- is a data organising infrastructure supporting discipline and research challenge needs .

Key Concept

Once that distinction forms we see two head offices:

•eResearch-Green infrastructure is research performance serving, it creates value from data and it can be taken broadly as an overall merit assigned research impact infrastructure

•eResearch-Gold infrastructure supports that research impact goal but must also serve other missions assigned to data:

- Stock of Knowledge – publication roles
- Asset Value – connecting data to create new knowledge domains and knowledge possibilities
- Innovation – data sharing across the sector and into other sectors

And a useful question is…..

Could the universities and publically funded research agencies signal at the highest level possible:

- *that they propose to focus research competitiveness around computation and the use of data [eResearch-Green],*
- *and to cooperate around access to data, and the cloud like application of data, including co-operating around the systems that borderless data relies on [eResearch-Gold]?*

Some things we can predict

Computing is going to ‘fill in’ between peak and the desktop

- Rising data volumes and data object sizes demand it
- Does it grow from peak, or from cloud, or is it a specialist mid range build out?
- We can have and probably want highly varied and different compute solutions

The data system has to meet many differentiated global community needs using as common a policy environment and as common an infrastructure as possible

- We don’t want every institution solving this problem a different way
- We have to become much more efficient at data (ie lower labour cost)
- eResearch should give DVCsR better data outcomes using less research FTEs

Overall research software is on a maturity trajectory we need to assist

- The management of the research software base needs to ‘get organised’

Provenance and Quality of Research needs attention

- Much as ANDS was set up a decade ago to create a focus on data, we might be advised to do the same thing for provenance and quality of research

Framework

Two Key Components

Head Office

National Computational Facilities

Head Office

Australian Research Data System

Two Supporting Components

Connectivity Operators

AARNet, AAF

Virtual Discipline Platforms

TBA

Which brings us to today

The eResearch capability is in a position to reframe its progress.

- Investment has created a suite of capabilities on which research dependency is increasing and from which research impact is arising;
- A one year funding period is available to operate existing infrastructures through 2016/17; and
- The prospect of a longer timeline investment from July 2017 is available following a road mapping and planning activity.

While there is a capital refresh challenge that is not resolved by the above, the situation presents as a critical opportunity:

a moment in time where a rethought future could be articulated;

because the means to embrace it exists.

Which brings us to today

All participants in the research sector will spend on eResearch simply because their researchers use computers, use data, use computers to use data, write software, import software, import data, publish data, or move data from one place to another. It is unavoidable.

How do we understand who spends on what and why?

Nothing can secure durability more easily than a long term commitment from a party that is itself sustainable. It follows that NCRIS investment in eResearch infrastructures will be more secure if institutions can build within, expand or adapt those infrastructures.

An NCRIS eResearch facility must be able to provide and tailor generic digital resources, provisioned at scale, to support many differentiated research activities to international standards.

If we put the actual activities under way aside, the question occurs, what is it that a regional deployment assists with from a national perspective?