NCRIS
National Research Infrastructure for Australia
An Australian Government Initiative

RDSI

Research Data Storage Infrastructure

The RDSI DaSH Subproject – the complementary services needed to make RDSI a success

RDS

Research Data Services

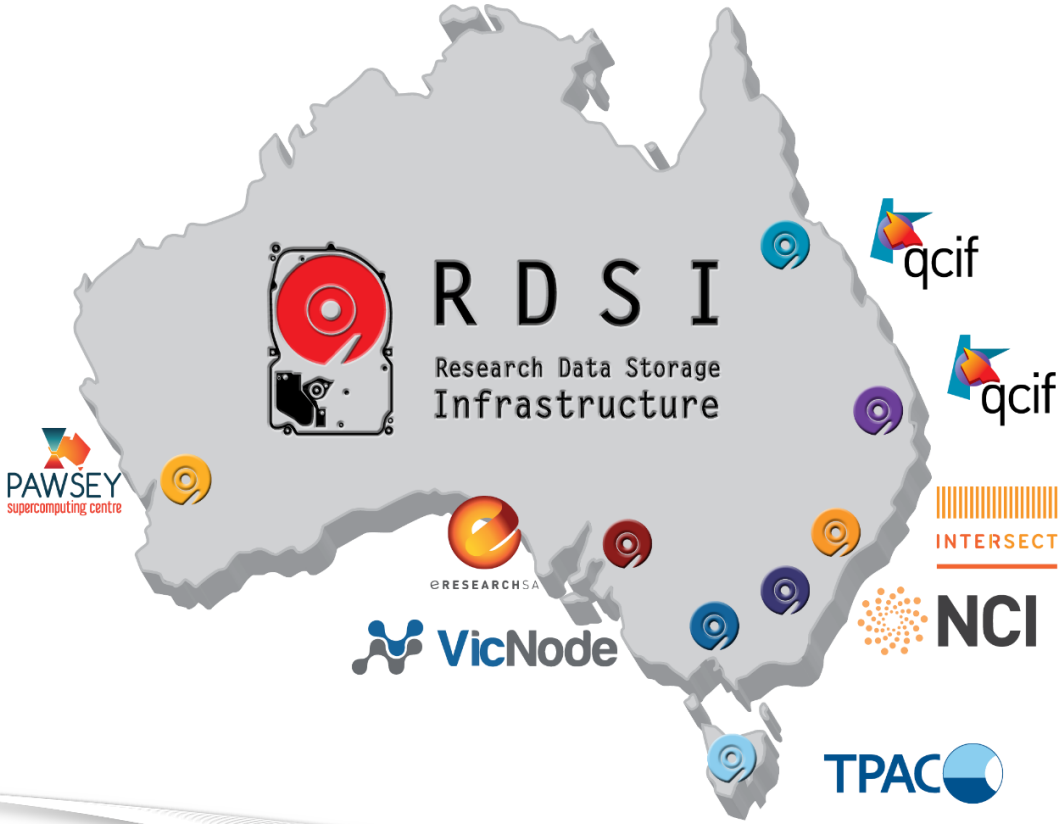
Life after RDSI

RDSI
Research Data Storage Infrastructure

A quick recap of RDSI

- Federally Funded Project
- Created storage infrastructure for nationally significant research data
- Data is accessible, shareable, discoverable and re-usable

RDSI funded Nodes



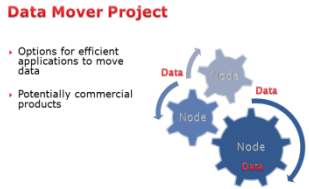
RDSI Nodes

- 6 Primary Nodes

- 2 Additional Nodes

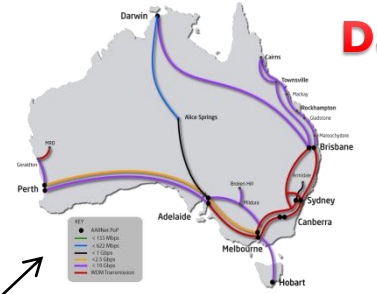
DaSh Programme

Data Mover

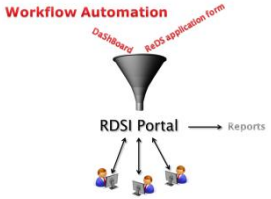


Test Platform

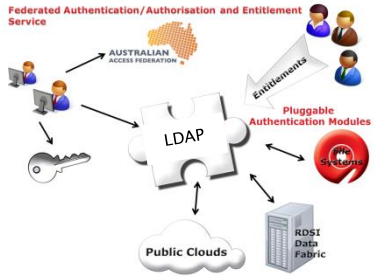
DaShNet



Workflow Automation

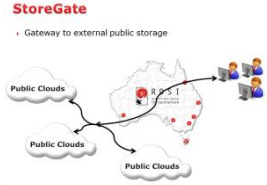


Identity



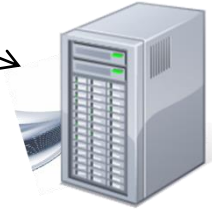
DaSh Components

Public Cloud Access



Security

Data Access Methods



RDSI Research Data Storage Infrastructure

Key DaSh Programme outcomes

ARMS

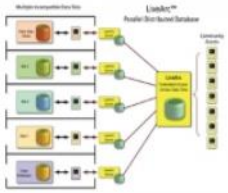
Data Sharing Network

Security

Identity

Authorisation Management

LiveArc



Reporting

High speed network powered by **AARNet**

DaShNet

- Reliable high speed network
- Built on the AARNet backbone
- Access to a public cloud infrastructure
- Science DMZ

RDSI
Research Data Storage Infrastructure

DaShNet

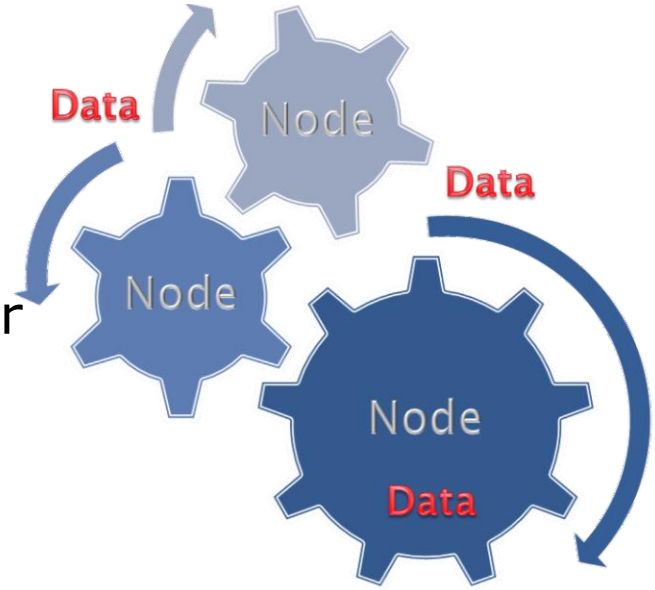
Fast network between nodes running on AARNet

Data Mover

- Large scale data movement is more than fat network pipes
- Replication between nodes at large scale needs efficient transfer mechanisms
- There are security challenges!

Solution

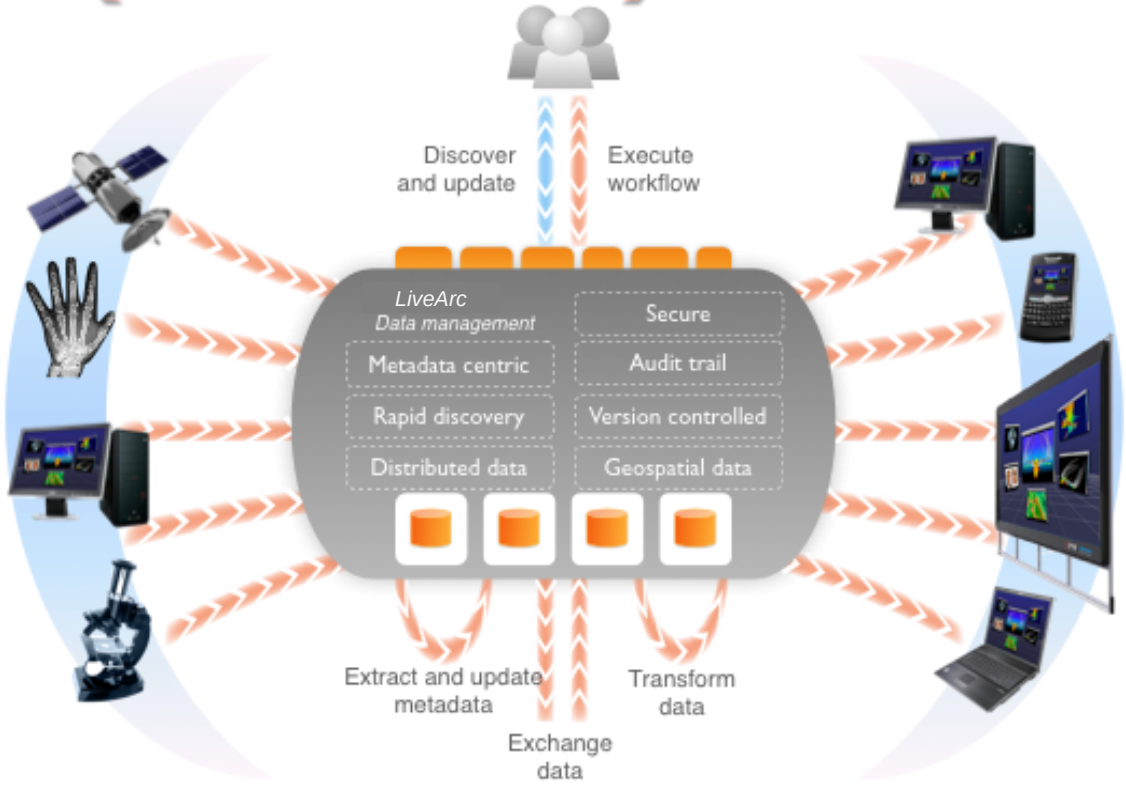
- Dedicated, high-performance Data Transfer Nodes (DTNs) based on Science DMZ from Energy Science Network
- Dedicated performance management
- Aspera software



Public Cloud

- Volume aggregation through CAUDIT/Test Platform
- Removing barriers to use
  - AWS IAM integration
  - Direct Connect
  - No egress charges

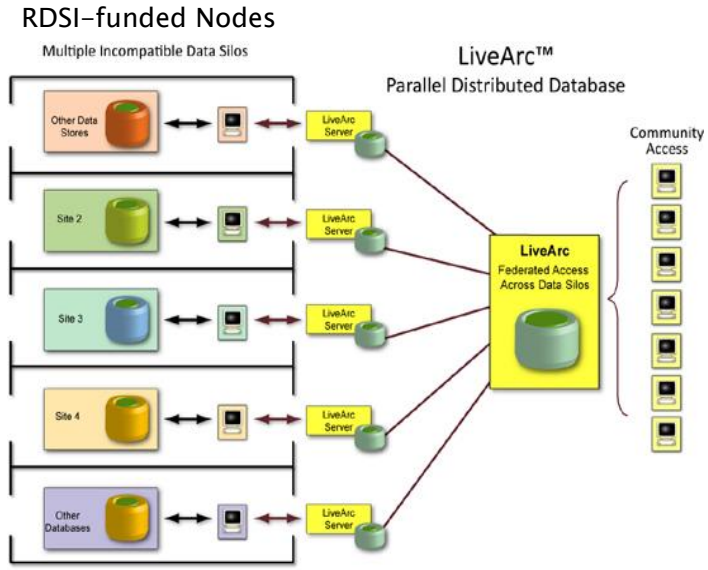
LiveArc (Mediaflux)



RDSI
Research Data Storage Infrastructure

LiveArc (Mediaflux)

- Allows users to directly organise, manipulate and search large amounts of data
- Ingested, data is immediately available for search and discovery
- Data easily and rapidly located
- Research data repositories at the Nodes can be federated with queries presented to the researcher as if only one repository had been queried



ARMS

- Allocation Request Management System

- Automated tool for merit-allocated storage assessment and allocation workflow

Security

- Developed recommendations on information security policy and best practices to RDSI-funded Nodes
- Outcomes
  - Assist in establishing a **robust security architecture for Nodes**
  - **Benefit Australian Researchers** and their international colleagues

Reporting and monitoring

Nodes report

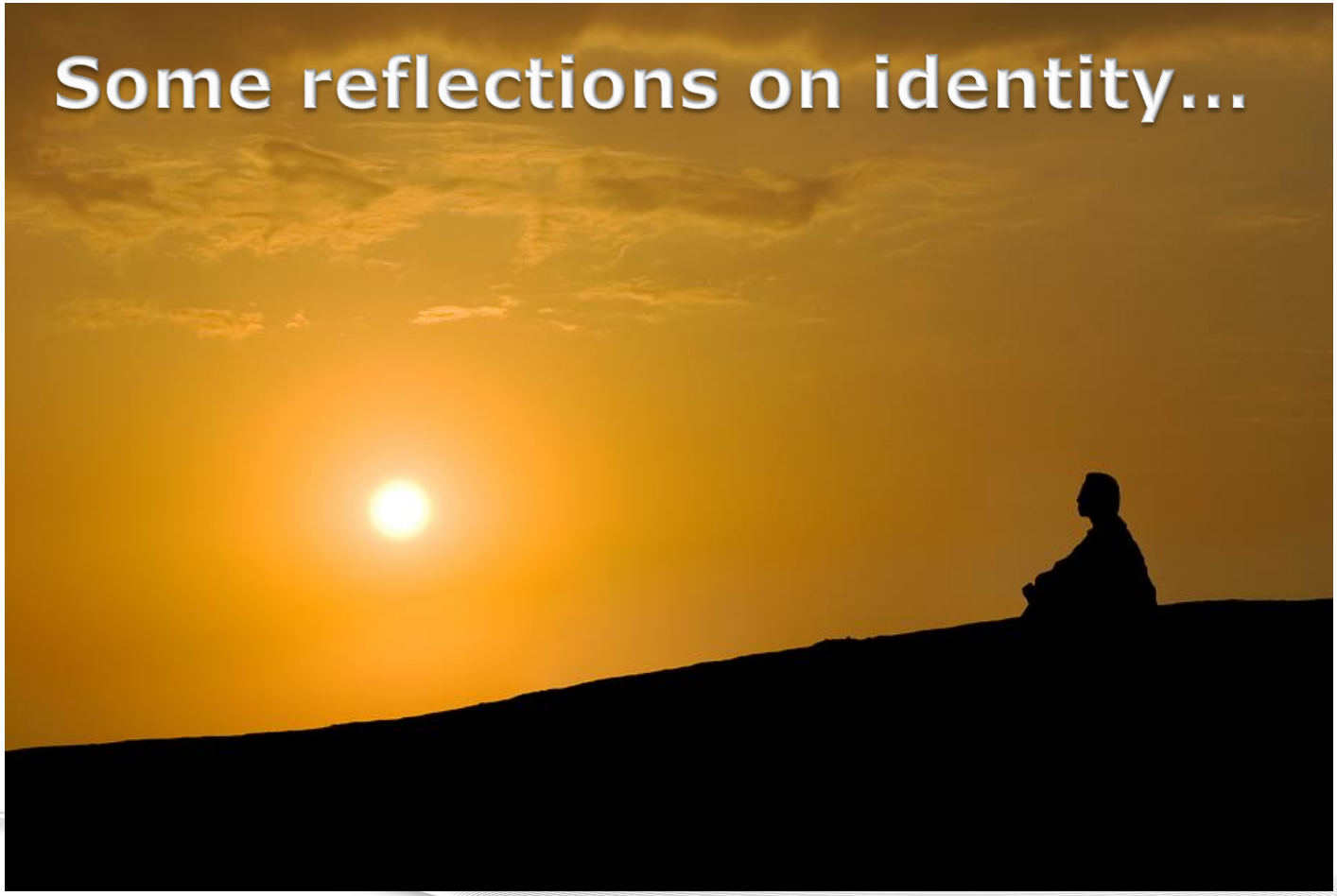
- On their availability to researchers
- Collections allocated by Node's merit allocation committees
  - Over 53 PB of collections have been approved

Identity and Authorisation

- Created a Researcher Identity Framework
- In partnership with:

AUSTRALIAN ACCESS FEDERATION

- Federated access to data


Some reflections on identity...

Non web scenarios

- Current Federation Technologies are based on SAML
- They don’t easily support Non Web applications
  - ECP provides some options
  - RDSI funded Livearc (Mediaflux) integration with SAML/ECP
  - Moonshot (based on radius) is nearing a point of some promise but still 1-2 years away from mainstream use

You need an identity federation!

- You need this from the beginning
- All researchers need to have federated ID’s
- Very difficult to retrofit it afterwards

Other researchers

- Your researchers WILL want to collaborate with other university researchers in other countries

- An international inter-federation initiative – Likely to be a number of years before benefits are realised

Non university researchers

- Social ID's is one alternative pathway
- You need to be flexible…

Access Management

- Researchers will give you highly complex scenarios for access management
- 90% of them can be distilled down to:
  - Administrator
  - Owner
  - Read/Write
  - Read Only
- Don’t allow yourself to be sucked into a black hole with complex scenarios

Important Partnerships

- Your IT Director/CIO is a key partner
- They often own the institutional identity infrastructure
- You need them onside

NCRIS
National Research
Infrastructure for Australia
An Australian Government Initiative

RDSI

RDS

Research Data Services

A federally funded project
which builds on RDSI

RDSI
Research Data Storage
Infrastructure

RDS Key Activities

- Maturing sustainable data services that deliver research outcomes and impact
- Quality, sustainable and co-ordinated operations

Research Areas of Focus

Climate and Weather Science

NCI TPAC

Life Sciences (Genomics)

qcif eRESEARCHSA INTERSECT VicNode

Medical and Health Sciences

INTERSECT eRESEARCHSA VicNode

Image Publishing

VicNode INTERSECT qcif ivec.org iVEC

Access to Data for Culture and Community Research

eRESEARCHSA INTERSECT qcif VicNode

Astronomy

ivec.org iVEC NCI

Terrestrial Systems Research

qcif NCI eRESEARCHSA qcif

Marine Science

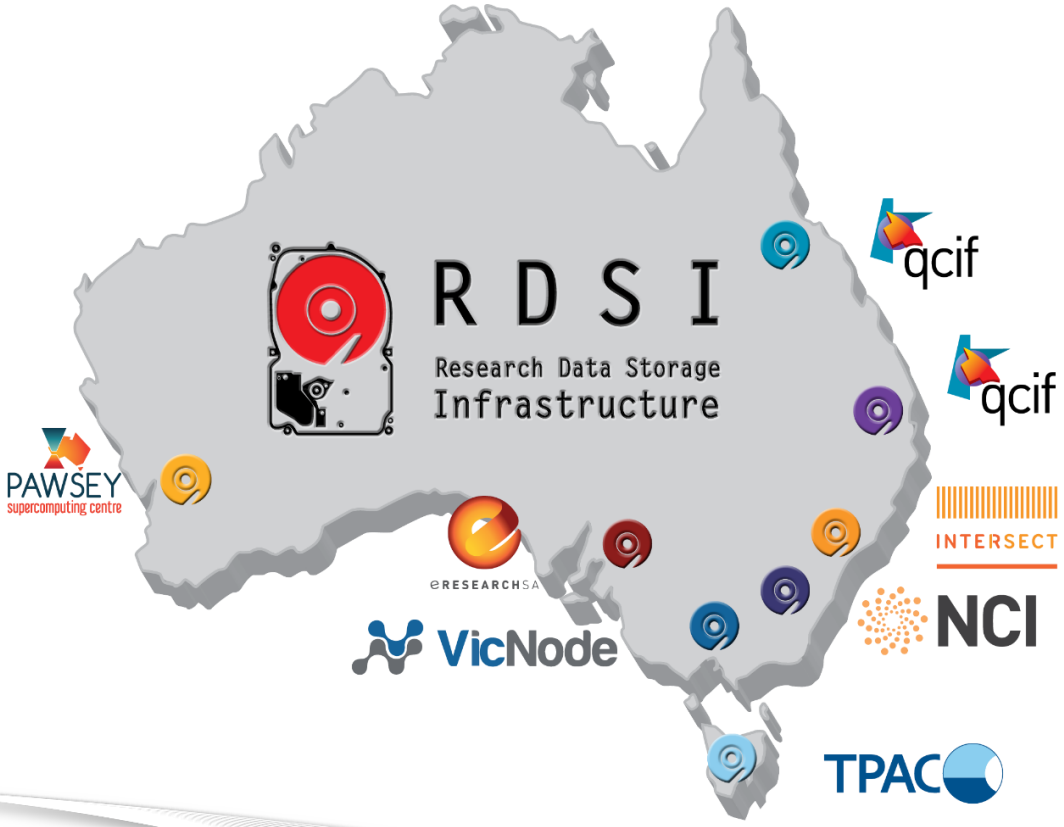
TPAC TASMANIAN PARTNERSHIP FOR ADVANCED COMPUTING ivec.org iVEC qcif

Geoscience

ivec.org iVEC

RDSI
Research Data Storage Infrastructure

RDSI funded Nodes



Quality, sustainable and co-ordinated operations

- Project Co-ordination
- Maintaining Quality, Security and Access
- Data Services Operations

RDS Project Dates

- Commencement date - December 2014
- Project end date – June 2016

Questions?

Richard Northam
Director
Research Data Services Project
richard.northam@uq.edu.au