



Australian Government

**Department of Industry, Innovation, Climate Change,
Science, Research and Tertiary Education**

DaShNet: Foundation for the Australian Research Data Cloud

Peter Elford, RDSI Network Programme Manager

p.elford@rdsi.uq.edu.au

3rd July, 2013



R D S I
Research Data Storage
Infrastructure



- ▶ Research Data Storage Infrastructure
- ▶ \$50M Federal government investment
 - Education Investment Fund (EIF)
 - Super-Science initiative (eResearch)
- ▶ Establish a nationally distributed federation of Nodes, to provide support for the capture, management, access and re-use of nationally valued research data collections



RDSI Nodes

- ▶ Partnerships or entities established by the universities in each state to deliver research-related ICT services
 - Selected via RFP
- ▶ Nodes are hosted in new/enhanced data centres
 - 6 Primary Nodes – larger, direct access to AARNet backbone
 - 2 Additional Nodes – smaller, shared access to AARNet backbone
- ▶ Nodes are expected to be sustainable entities beyond the RDSI project
 - Operational support funded through membership, subscriptions, fee-for-services, etc.

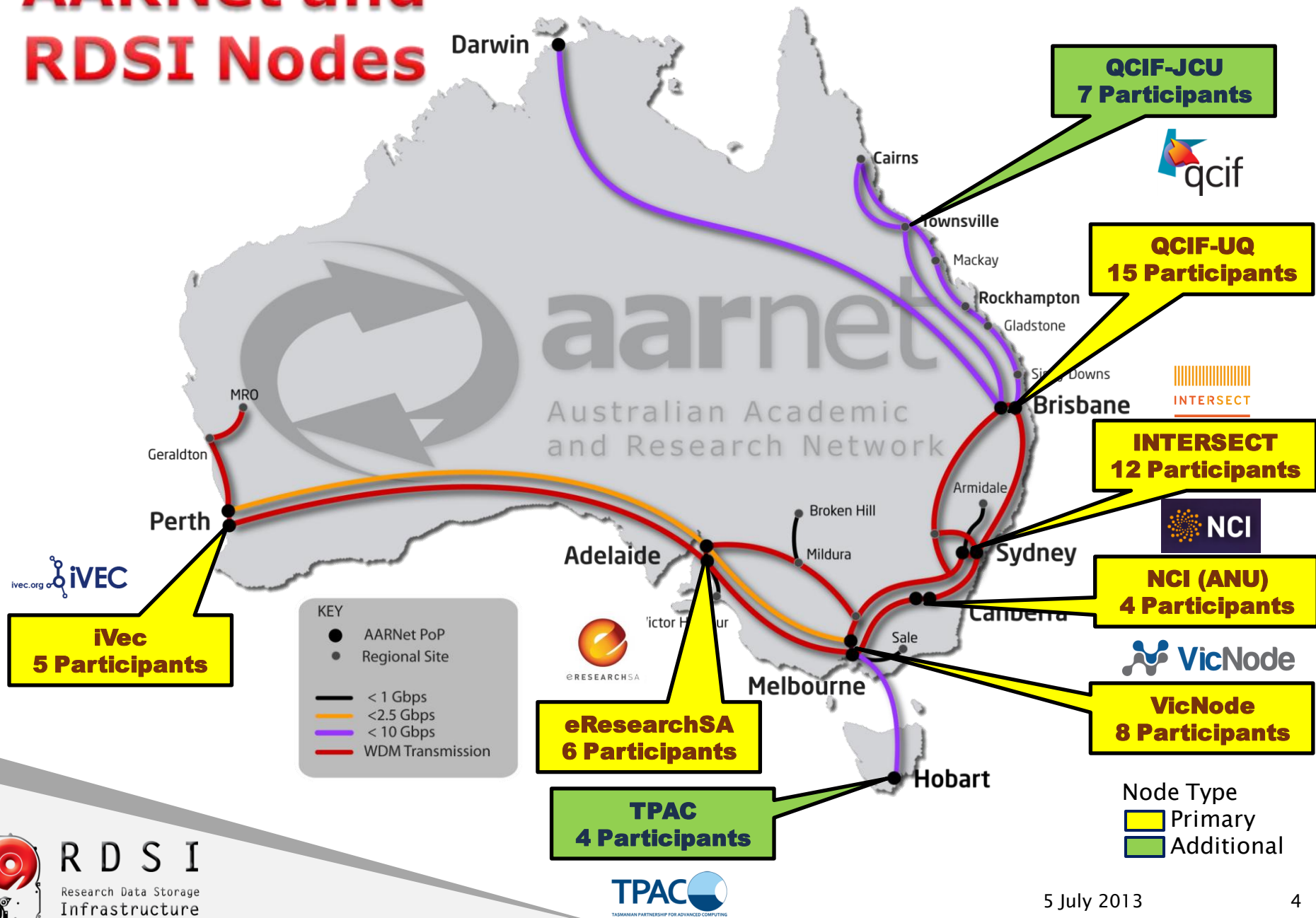


eRESEARCHSA



R D S I
Research Data Storage
Infrastructure

AARNet and RDSI Nodes



RDSI Components

Nodes

- Identify, strengthen and develop research data centres able to hold and process high data volumes

ReDS Research Data Service

- Identify research data holdings of lasting value and importance and contribute funding to their development at the most appropriate Node(s).

DaSh Data Sharing

- Build capability to support the sharing and re-use of research data. Five themes:
 - Networks
 - Data Access
 - Identity
 - Workflow
 - NodeZero

VePa Vendor Panel

- Establish a panel of vendors for RDSI Nodes and sector



DaShNet

- ▶ Network Part of RDSI Data Sharing (DaSh)
- ▶ Funded by National Research Network (NRN)
 - Administered by UniSA
 - One of nine NRN projects
- ▶ Single proposal to NRN
 - RDSI (lead), AARNet, Nodes

- ▶ **NOT YET FORMALLY APPROVED**
 - **Remainder of talk subject to change**

1. Adelaide–Perth
2. Sydney Metro
3. Sunshine Backbone
4. Brisbane Metro
5. Tasmania Regional
6. VERN
7. WA Metro
8. MRO
9. **National Data Centre Interconnection**



High Level Objectives for DaShNet

- ▶ Like RDSI itself, NRN is EIF funding
 - Infrastructure only
- ▶ Enable RDSI
- ▶ Provide National outcomes
 - Consistency
 - Big picture approach (part of overall NRN)
- ▶ Step change in network capability
- ▶ Minimise future OPEX
 - Capabilities funded by RDSI are delivered and **sustained** through the Nodes



Outcomes, Elements

- ▶ High speed broadband **access to** Nodes
 - Dedicated dual 10G access to AARNet
- ▶ High speed broadband **between** Nodes
 - Dedicated dual 10G VPN over AARNet
- ▶ Data Centre interconnections optimised for data-intensive research
 - Science DMZ
- ▶ Better network performance management instrumentation
 - Science DMZ
- ▶ Access to public cloud infrastructure
 - Fibre to selected data centres
- ▶ Access to external data sets
 - Fibre to selected data sources to allow ingest to Nodes

Fibre

Network

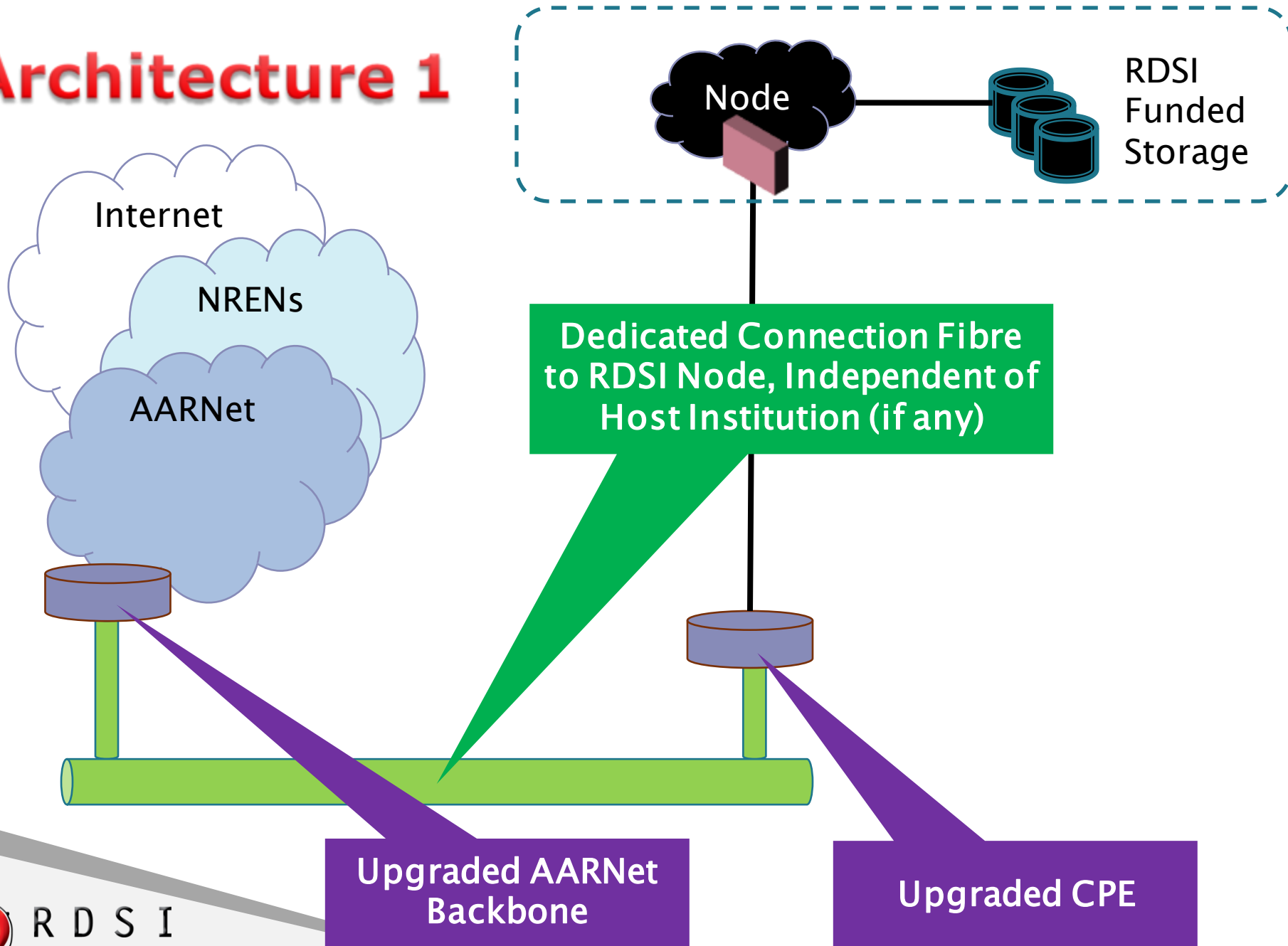
Other

1

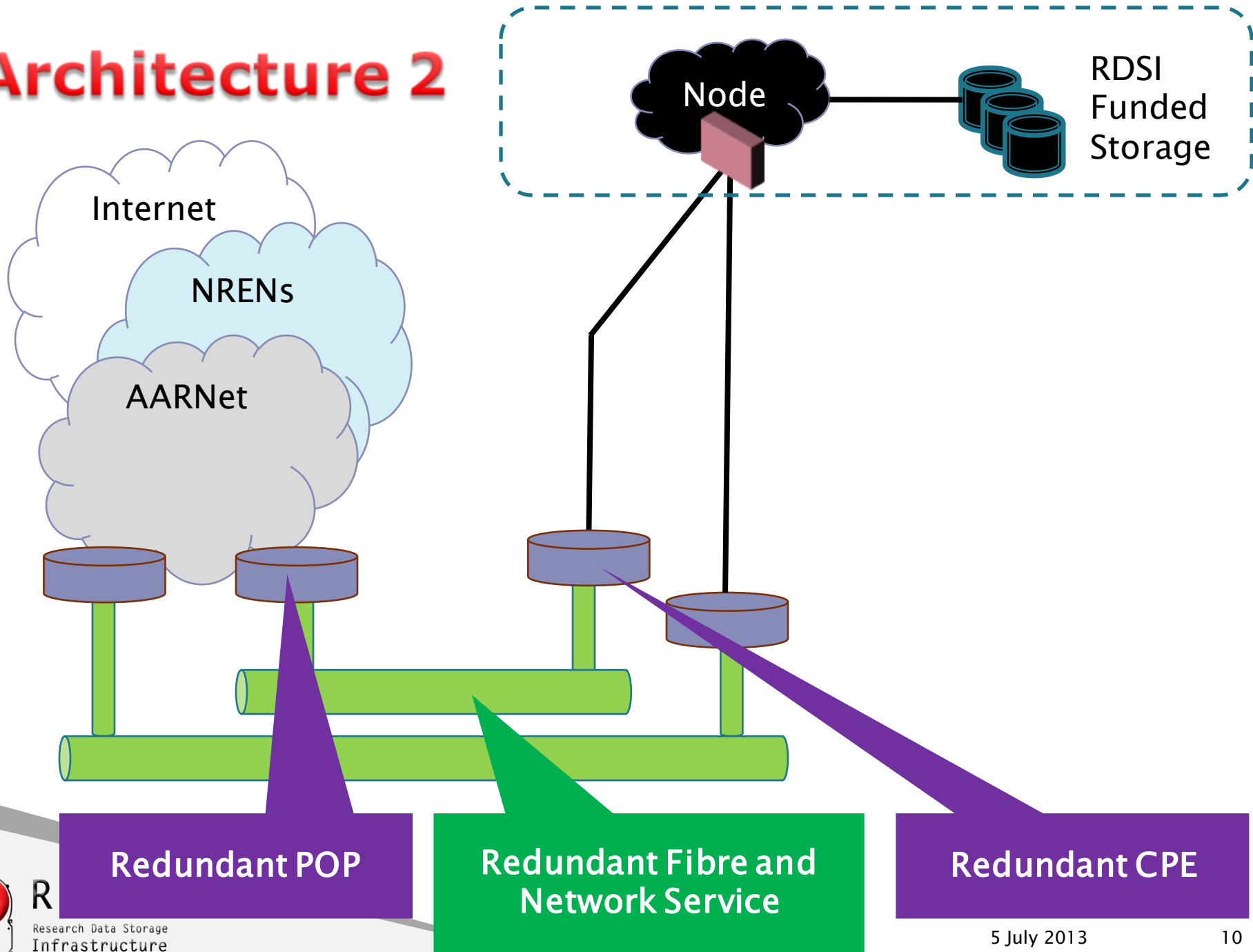
2



Architecture 1



Architecture 2



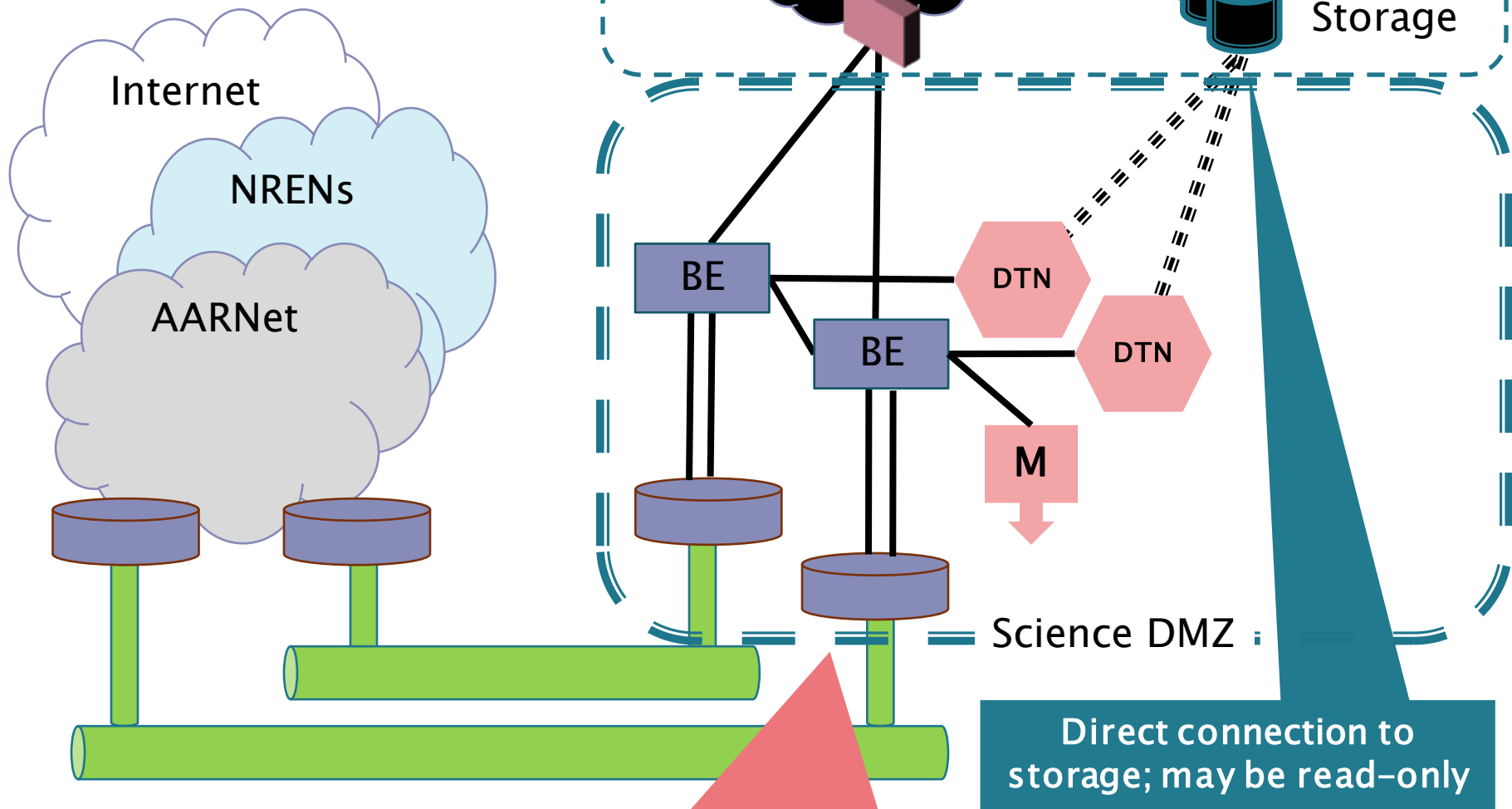
R

Research Data Storage
Infrastructure

5 July 2013

10

Architecture 3

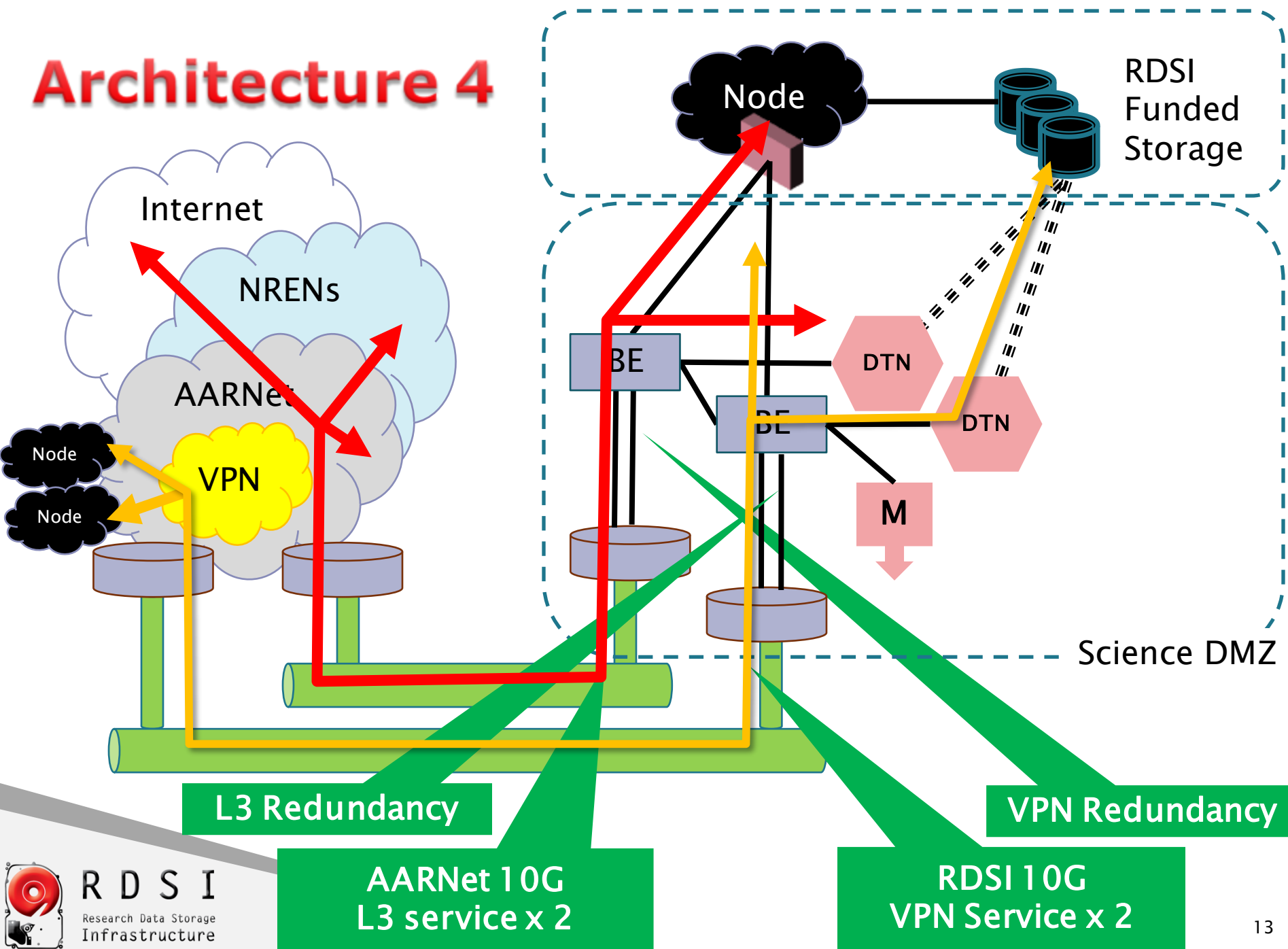


Why Science DMZ ?

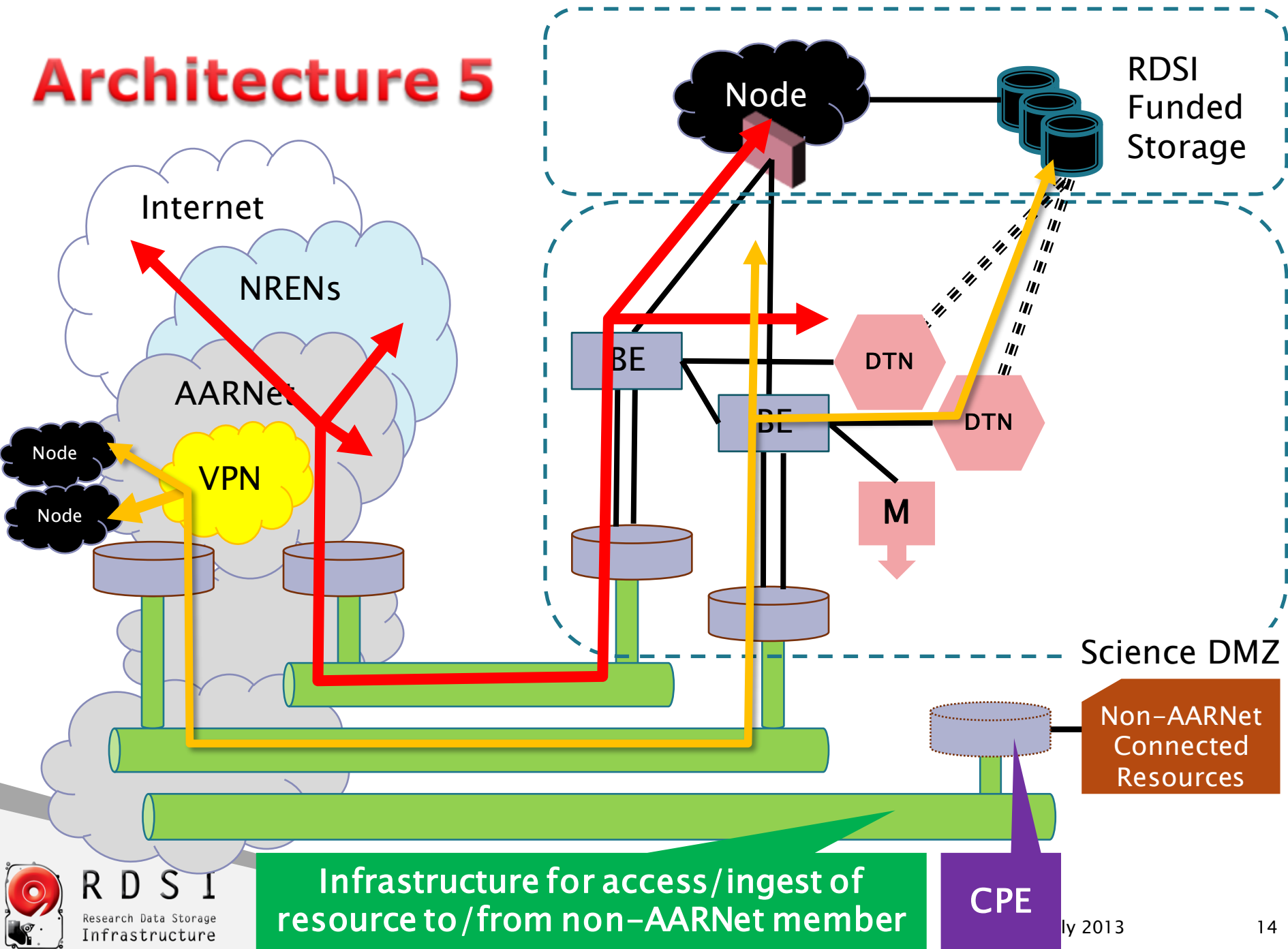
- ▶ Based on Energy Science Network (ESnet) Science DMZ (<http://fasterdata.es.net/science-dmz>)
 - Large scale data movement is more than fat network pipes ... **particularly in Australia**
- ▶ Three key components – **ALL REQUIRED**
 - Friction Free Network
 - Wire-speed/deep queue network devices, drop free, research-specific security policy (outside organisation perimeter)
 - Dedicated Data Transfer Nodes
 - Optimised for BULK file transfer over long network pipes
 - Performance Measurement Nodes
 - Validating network performance and value



Architecture 4



Architecture 5

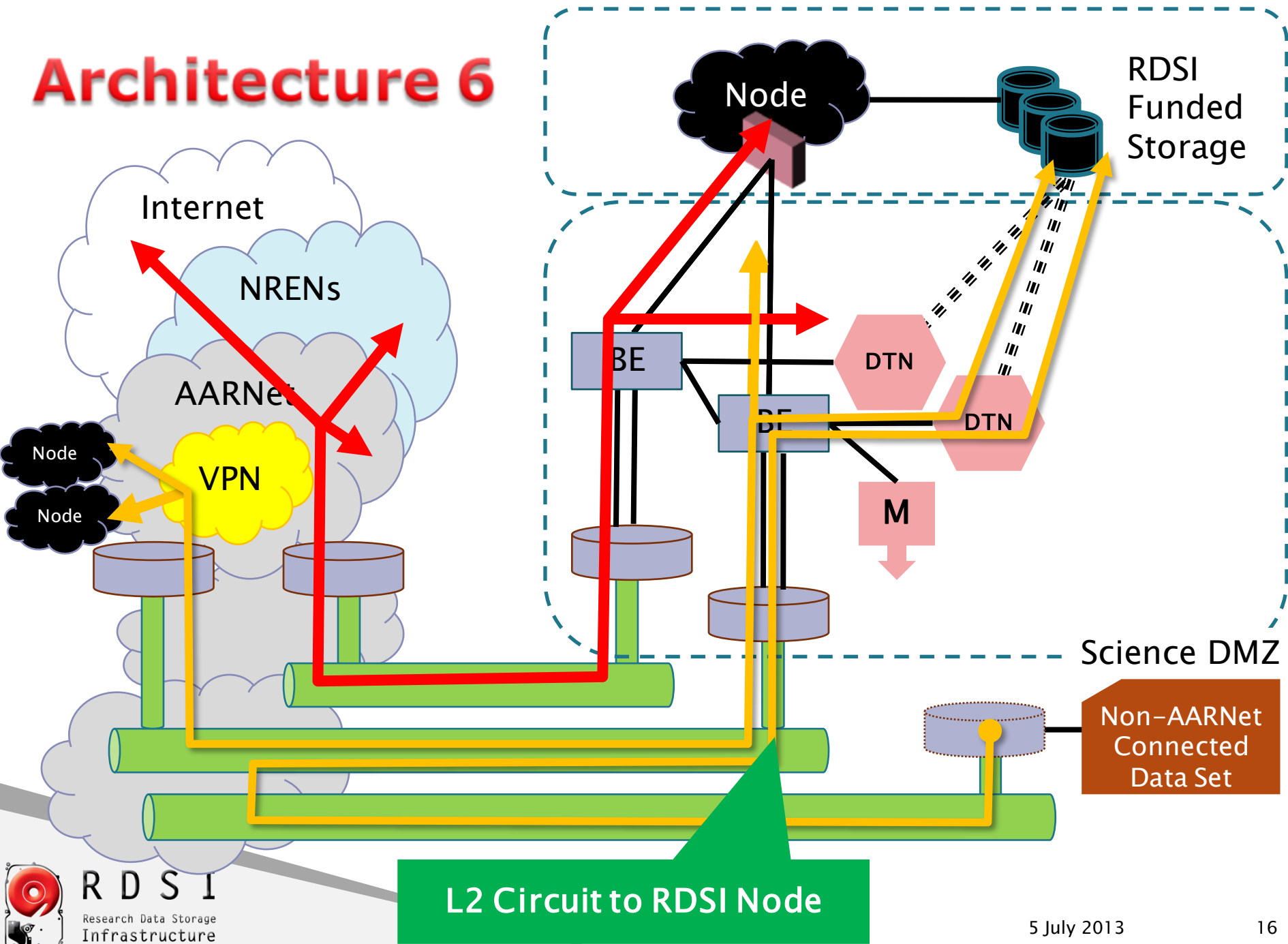


Non-Member Resources

- ▶ Commercial Providers of (Cloud) Services
 - Amazon Equinix, Sydney (**mostly done!**)
 - Others TBD (Rackspace, Dell, HP, ??)
- ▶ Data Sets within Government/Research
 - State government networks/agencies
 - Federal government networks/agencies
 - Potentially, leverage whole of government networks
- ▶ Range of options for connection outside standard AARNet membership
 - Peering (eg. Amazon, Google, etc.)
 - L2 connection to RDSI node



Architecture 6



DaShNet – Summary

- ▶ VERY significant network capacity for RDSI Nodes
 - Nodes will have more network capability than research-intensive universities (40G+)
 - Accelerates introduction of AARNet 4-based services
- ▶ New bulk data movement capability
 - Ability to easily utilise network capacity to support and deliver storage services
- ▶ Introduces the Science DMZ paradigm
 - A model for institutions to follow ?
- ▶ *Thanks to AARNet, RDSI Nodes, and NRN for their support in developing the DaShNet **proposal***
- ▶ **DaShNet NOT YET FORMALLY APPROVED**

Overview of the Science DMZ

Brian Tierney, ESNet – Thursday 10:50, Marquis



Questions ?



RDSI

Research Data Storage
Infrastructure