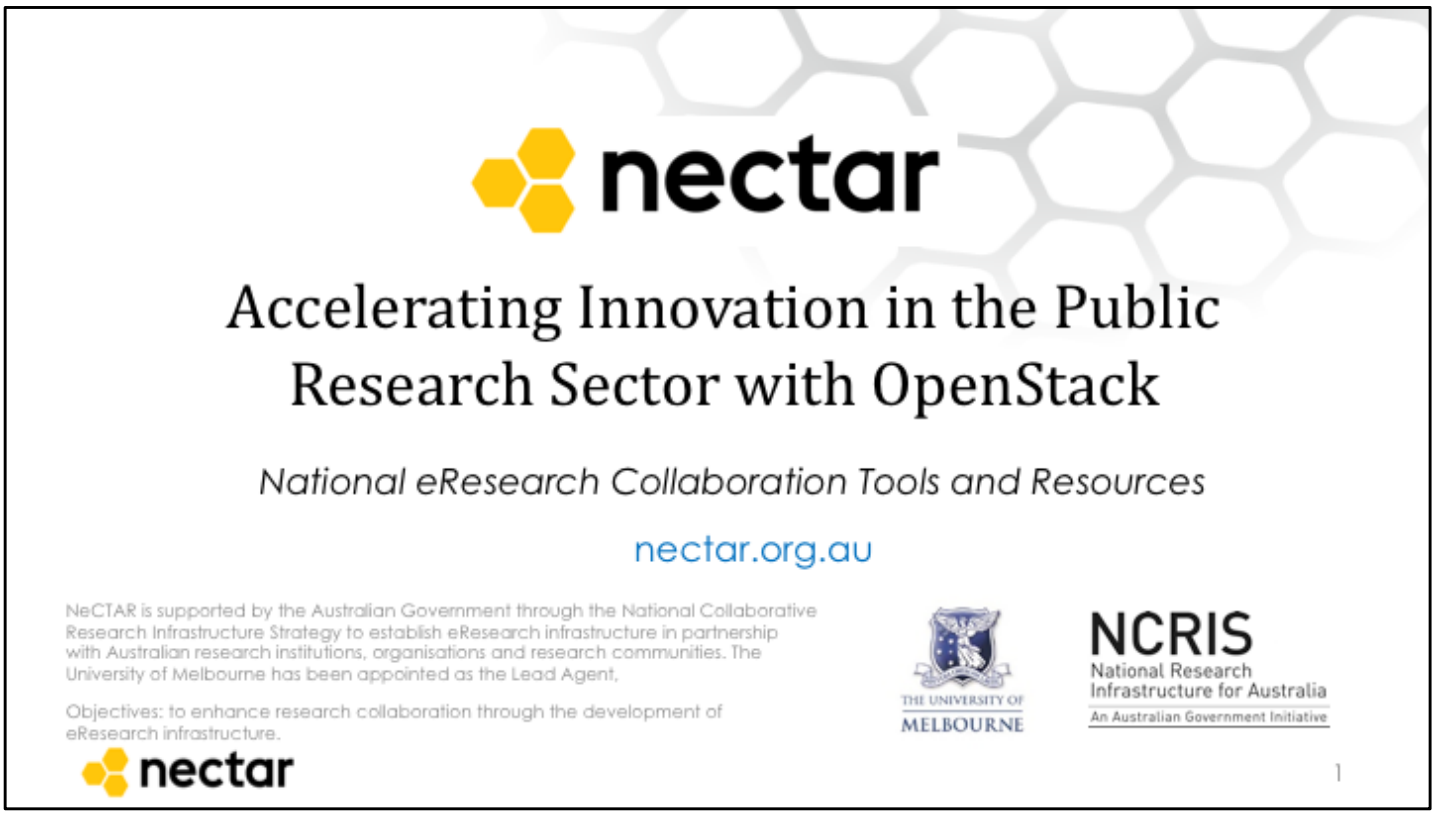


Key topics
- the evolution of the cloud as a technical platform
- the 'nodes', how they interoperate, and how they differentiate and add value to their research communities
- And what next?

Australian eResearch Infrastructure

EIF Super Science eResearch Investments - 2009-2014:

Shared Data:

▪ Australian National Data Service (ANDS)	AU$48M

Research Apps, Collaboration, Cloud

▪ ***NeCTAR***	**AU$47M**

Data Storage

▪ Research Data Storage Infrastructure (RDSI)	AU$50M

High Performance Computing

▪ National Computational Initiative (NCI)	AU$50M
▪ Pawsey Centre	AU$80M

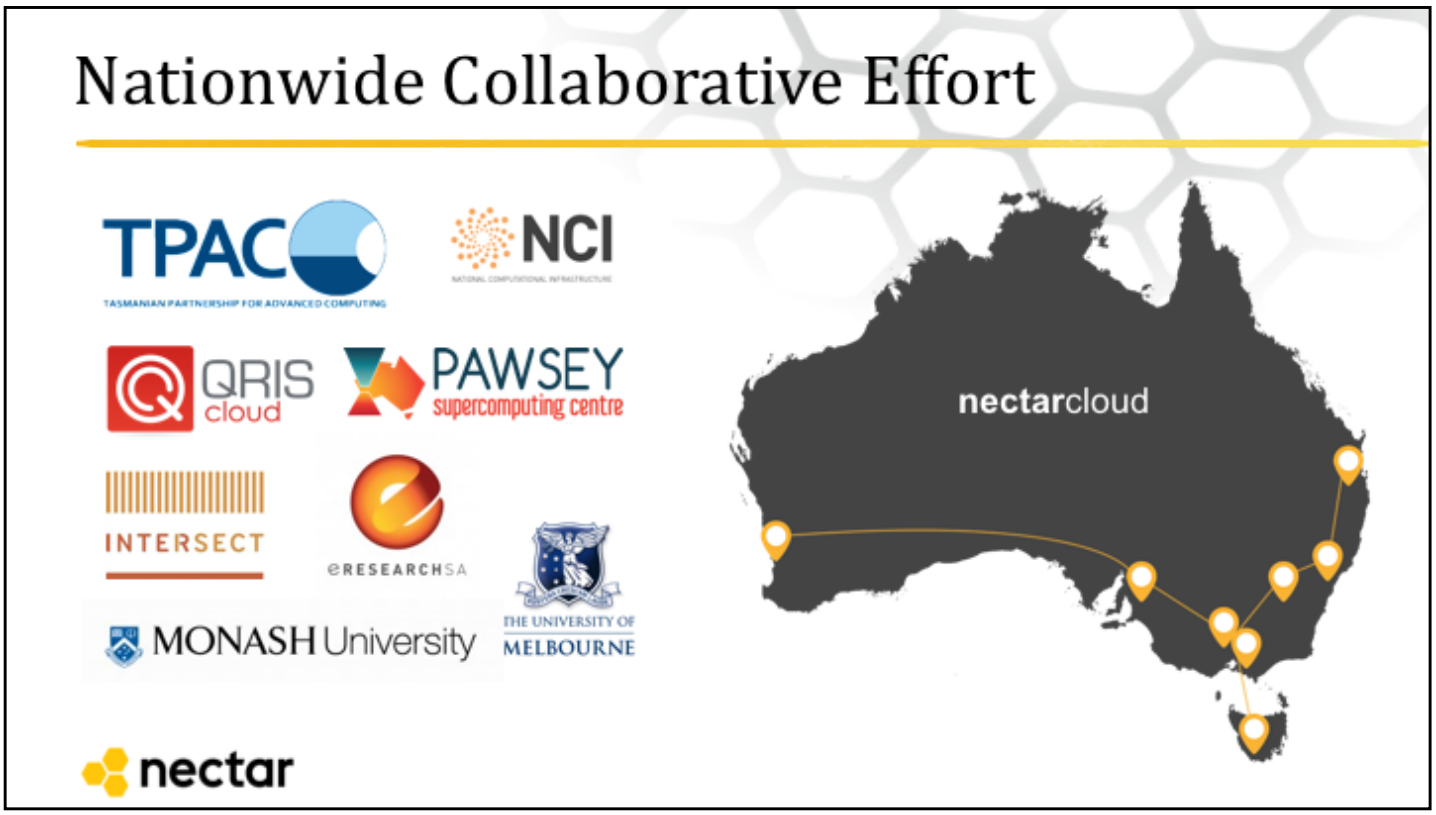
Networks

▪ National Research Network (NRN)	AU$37M

nectar

The nectar program was actually part of a broad set of eResearch investments over 5 years in Australia. These eResearch investments totaled around $300M over a period of 5 years. Total EIF SuperScience investment across a range of domains was $1.2B, with around $300M invested through the Super Science program.

At the time, NeCTAR and RDS (compute + data) were separated. In hindsite this wasn't such a great idea..



So I'll start with - where are we today?

The NeCTAR Research Cloud is a single integrated cloud operated by 8 national partners and supporting over **10000 research** user registrations across Australia.

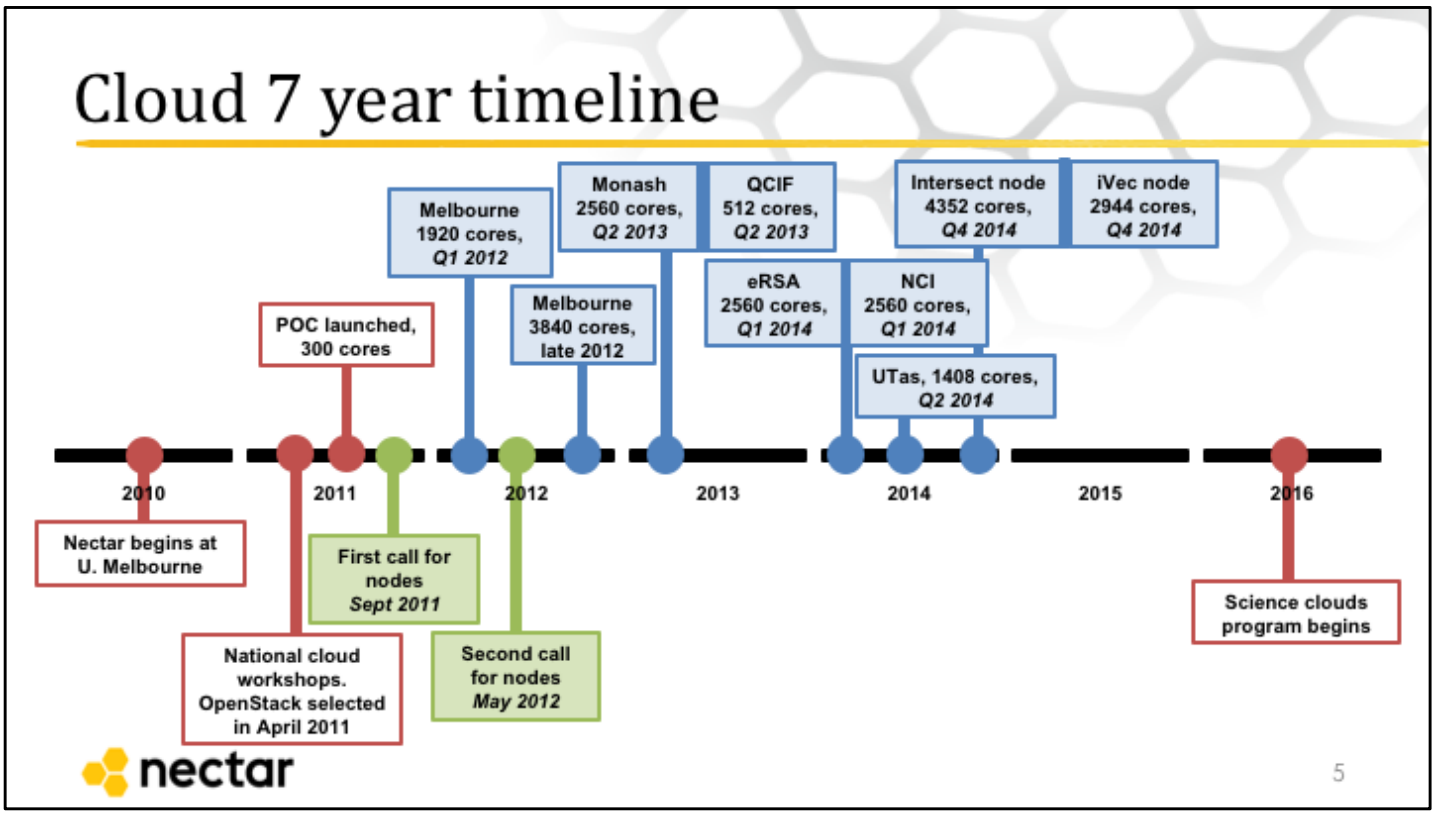
The thing to note here is that these are all quite different organisations. Some are universities, some are supercomputing centres (with a discipline twist), …

Nationwide Collaborative Effort

The NeCTAR Research Cloud is a single integrated cloud operated by 8 national partners and supporting over **10,000 research** user registrations across Australia. This innovative infrastructure is **reducing barriers to collaboration and knowledge sharing across institutional, state and national boundaries.**

nectar

This infrastructure was designed to **reduce barriers to collaboration and knowledge sharing across institutional, state and national boundaries.** The fact that we can take servers and data and ship them across the country, where you can access all this infrastructure using a single API and a single Dashboard, are all in support of these goals.



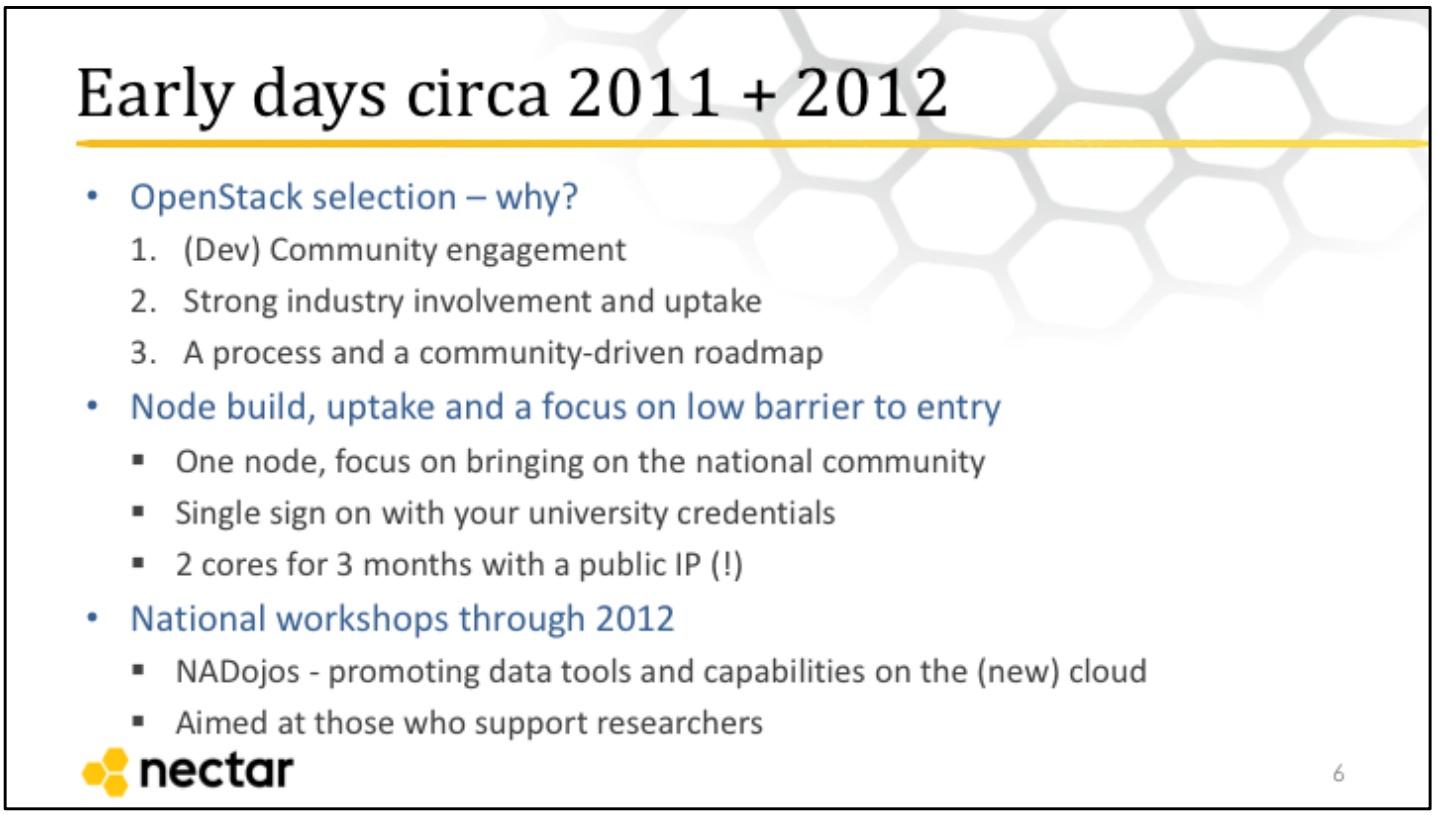
That's where we are at now but there's quite a bit of history stretching all the way back to 2010 which got us to where we are today, and I think are some broad themes and features of the approach very much worth appreciating what's happened here.

From the very early days, there was a rallying of the community even way back in 2010 to start getting everyone on board. In late 2010 there was a roadshow to introduce Australian University's to the nectar program, the inaugural director and they're v. small team .

The technical experts from around the country were invited to a workshop in Feb 2011 to talk about what and how. How would we build the cloud? And using what?

There were two node calls, inviting organisations to propose cloud nodes.

The overlay that isn't show here is that of the eResearch tools and Virtual laboratory program – which is a significant outcome of the Nectar program…



There were some specific activities that happened through 2011 and 2012, these were designed to establish the cloud and the community early on:

OpenStack selection : I mentioned getting the technical community together early on to decide on how to select a cloud platform. For cost, licensing complexities, and so on, various commercial solutions were out.
Proof of Concept – through 2011, 300 core test cloud which was used as a Poc…

Build of the first node & uptake:
After the POC we had the challenge at Melbourne of building the first cloud node. I will say, some of the early challenges were internal. We were part of ITS which was traditionally an enterprise shop.

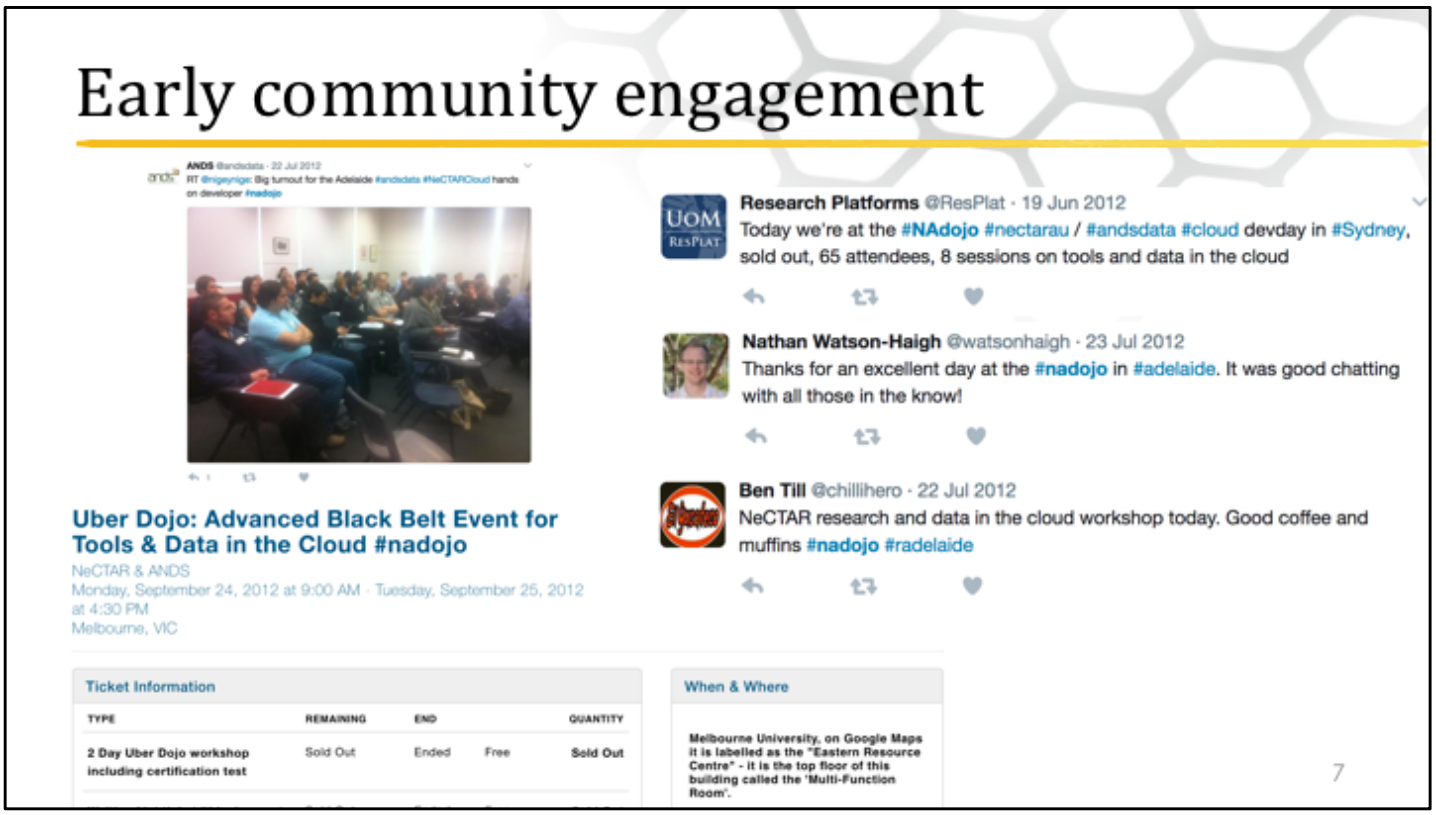
Where's the business case? Where's the architecture? We're just going to build it as we go – experiment with the infrastructure? What's does the service look like? We'll only know when people start using it. What's the SLA? We couldn't possibly use opensource technologies to run INFRASTRUCTURE. What do you mean you're going to give people a public IP address on University infrastructure? These were the conversations I would have continuously through the development of the first node, working to shield the technical team from the broader ITS.

However, it was made easier because we had the CIO on side, and he was the one

who vouched for us and was pivotal in bringing the Nectar project into Melbourne.

Early wins – lowering the barrier to entry – AAF login, 2 cores – this might shock you – but we had no T&C's,
At Melbourne was the first node, aimed to establish a groundswell of users across the country. What happened then was that we had researchers from multiple institutions utilising that (single node) cloud infrastructure as it was going to be about 14 months till another cloud node came online.

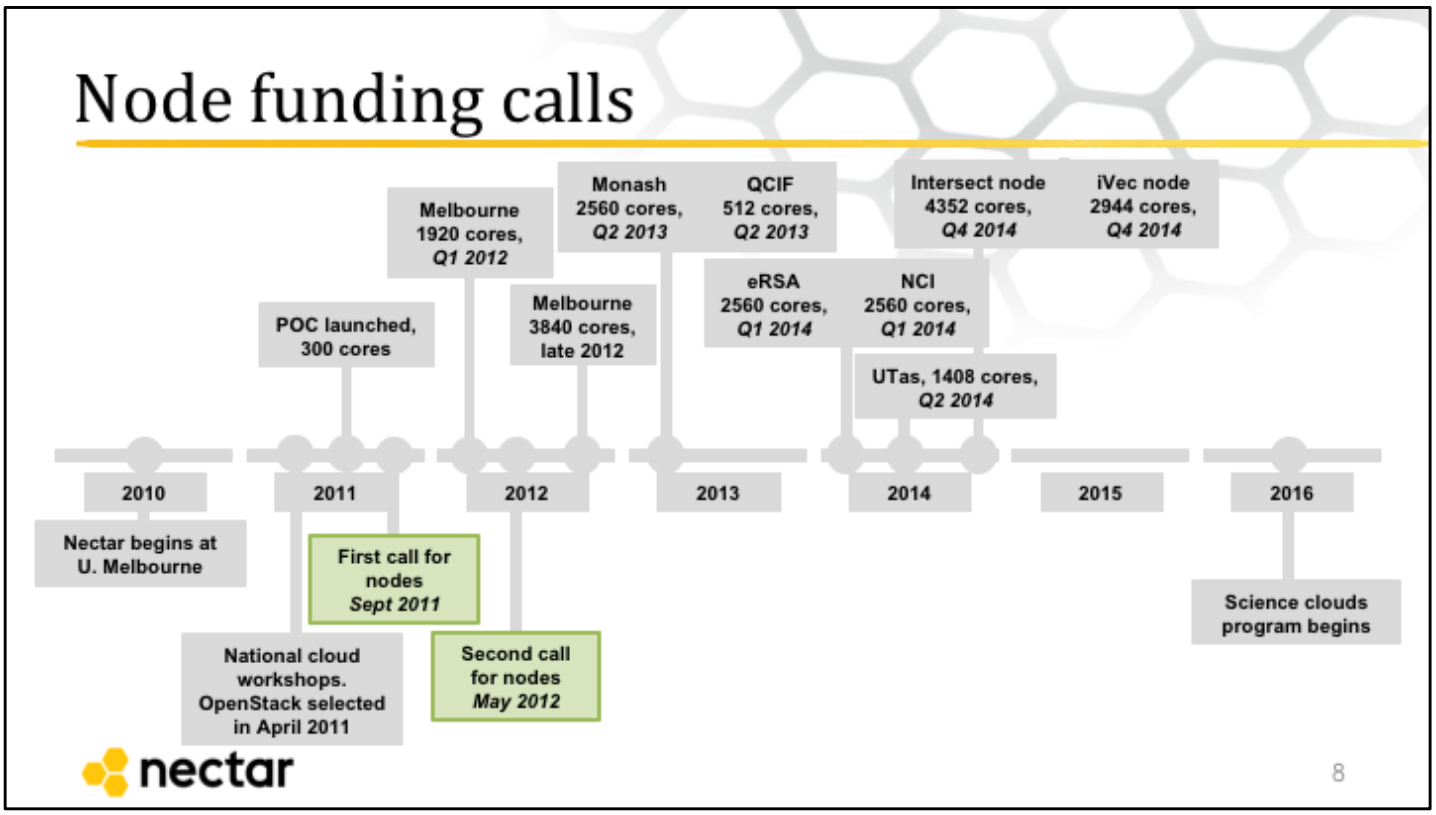
Next slide:
National NADojo workshops
- Early engagement
- ALSO 'Research cloud forum' @ Unimelb



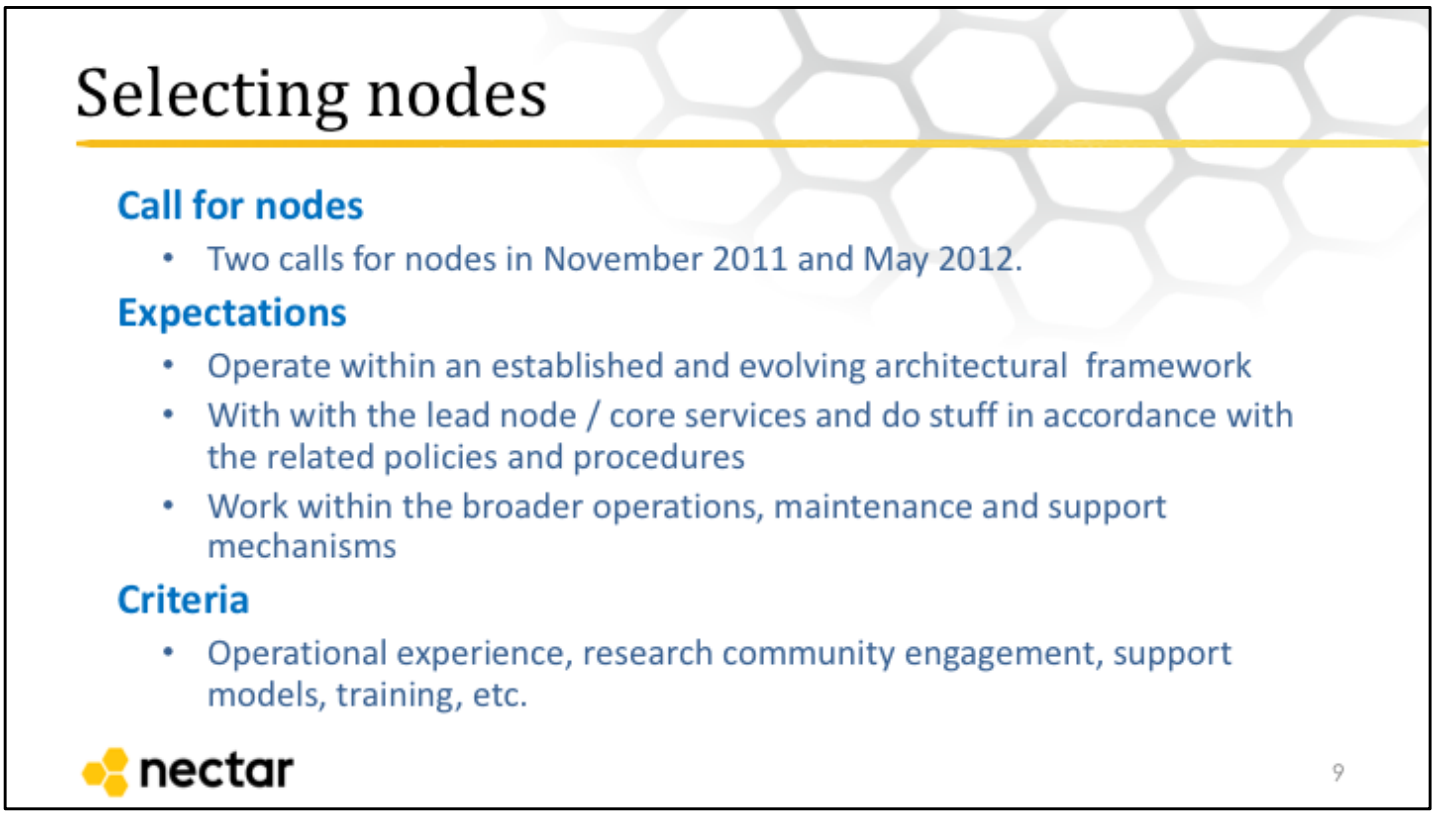
Early engagement
Focus on a national tour atop operating infrastruture
Dojo style training methodology, designed to ‘train the trainers’ – experts – technically proficient people who were already supporting researchers in eResearch
This literally involved members of staff from the University of Melbourne team going from capital city to capital city on an NADojo tour..

As we had the node at Melbourne, we also rad Research Cloud forums…

These early engagements were essential in building up a strong user community. Some of our most prominent supporters were users on Day 1, attended the forums, and are still very active users today.



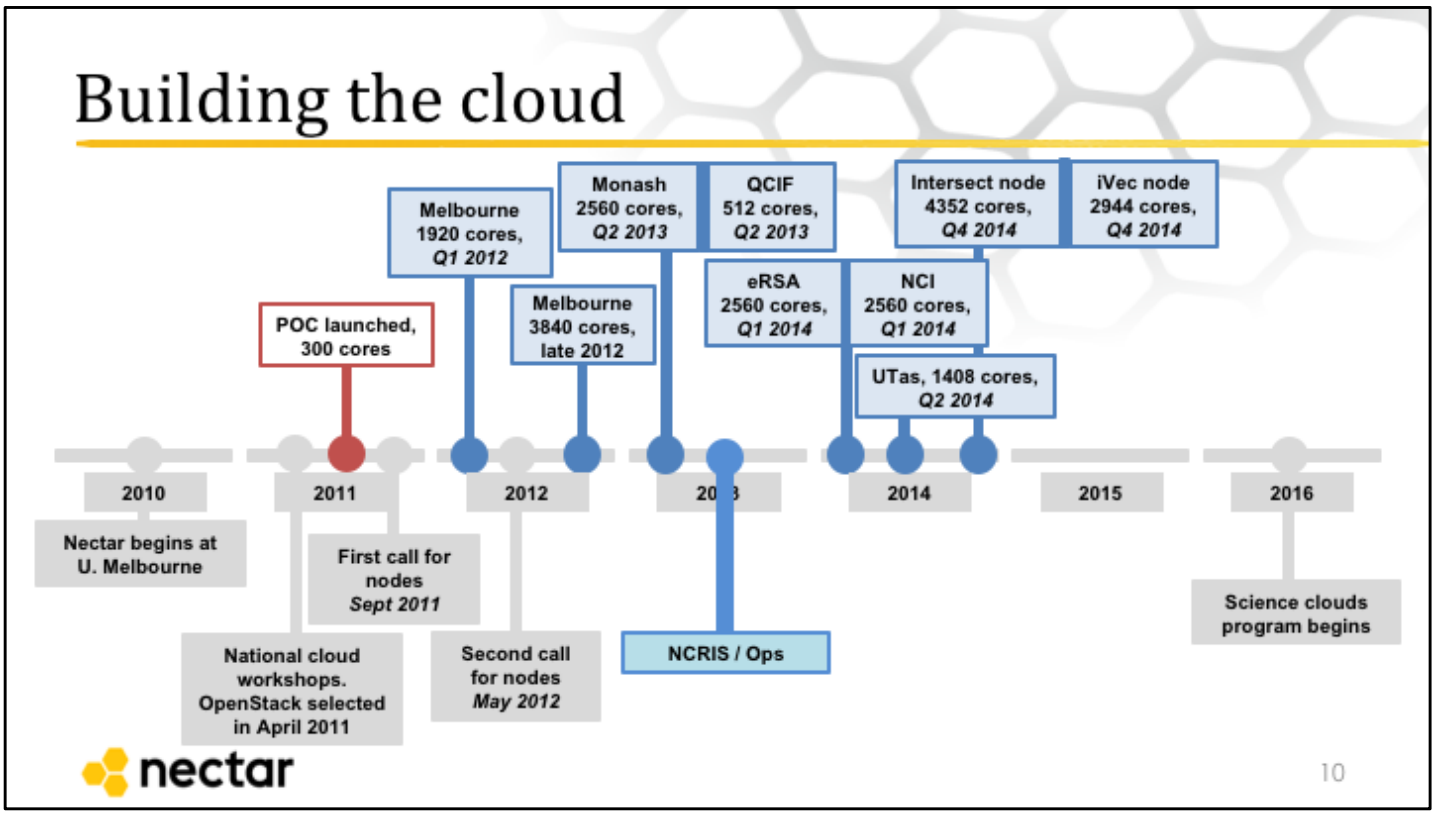
Around the time of the first node being launched… there was a formal request for proposal process to establish the nodes of the cloud. Operators, institutions and other affiliated organisations were invited to put in proposals for research cloud nodes.



Expectations - Note that all these things were evolving.

EIF funding – funded capital $, not operations, so each node had to do its own thing.

Expectations, the criteria…



… Build up of the cloud from the POC through the UoM doing the first node, to various funded nodes coming online. The total scale of the infrastructure is ~ 40K cores (funded mostly through NCRIS as well as institutional coinvestment).

The lead node idea started evolving as well.

In 2013, when NCRIS first came online, operations, maintenance and so on started to become supported. So the core services idea was developed further

Refining the approach, 2013+

"Work Packages" for Improved & Continued Operations:

Code	Title / Activity
WP1	Cloud Core Services – Lead: **Melbourne**
WP2	Monitoring and Reporting Services - Lead: **NCI**
WP3	Security Monitoring & Incident Response - Lead: **Intersect**
WP4	Quality Assurance of NeCTAR VMs - Lead: **Monash**
WP5	Continuous Improvement - Lead: **QCIF**
WP6	User Support – Lead: **QCIF**
WP7	Advanced Ecosystem Services - Lead: **Melbourne**
WP8	Cloud Allocations Policy & Improvements - Lead: **Monash**
WP11	National Server Program - Lead: **Melbourne**

3 operations streams
- Research Cloud Core Services
- User Support
- National Server Program

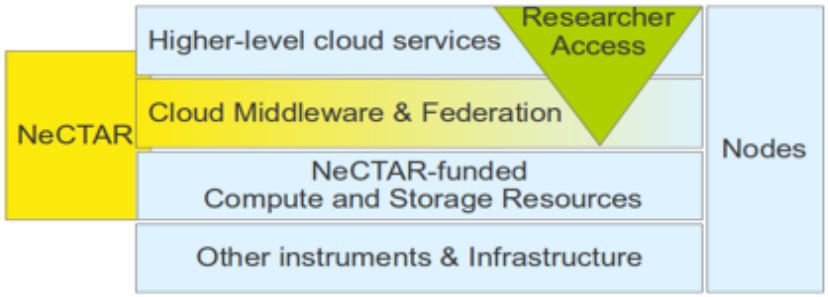
Ops improvements
OpenStack consistency across nodes
Reduced effort in upgrades
Improved core services uptime >99.895% first half 2015 and now regularly at 100%

NCRIS from 2013 brought in funding to support an uplift of capabilities ..

A single, federated infrastructure

An OpenStack based cloud infrastructure

- A single cloud deployed across 8 host organisations
  - Implemented using ***OpenStack Cells***



Nodes of the cloud are able to differentiate on a number of levels, while being part of a federation. Researchers will mainly use higher level services.

So what we ended up with was this….

So, with the middleware all strung to a central API endpoint, how different can nodes look?

Nectar OpenStack deployment

A single cloud dashboard (and API endpoint)

- ***OpenStack cells*** to support 8 regional sites:
  - Early adopters of Cells - *thanks RackSpace*
- Users can request a site – or *deploy anywhere.*

Global Object Storage federation (*Swift)*

- Providing data redundancy across sites

IaaS & platforms

- Standard & specialised instance configurations
- 'Golden images' and community contributed images
- Storage services (block, swift…)

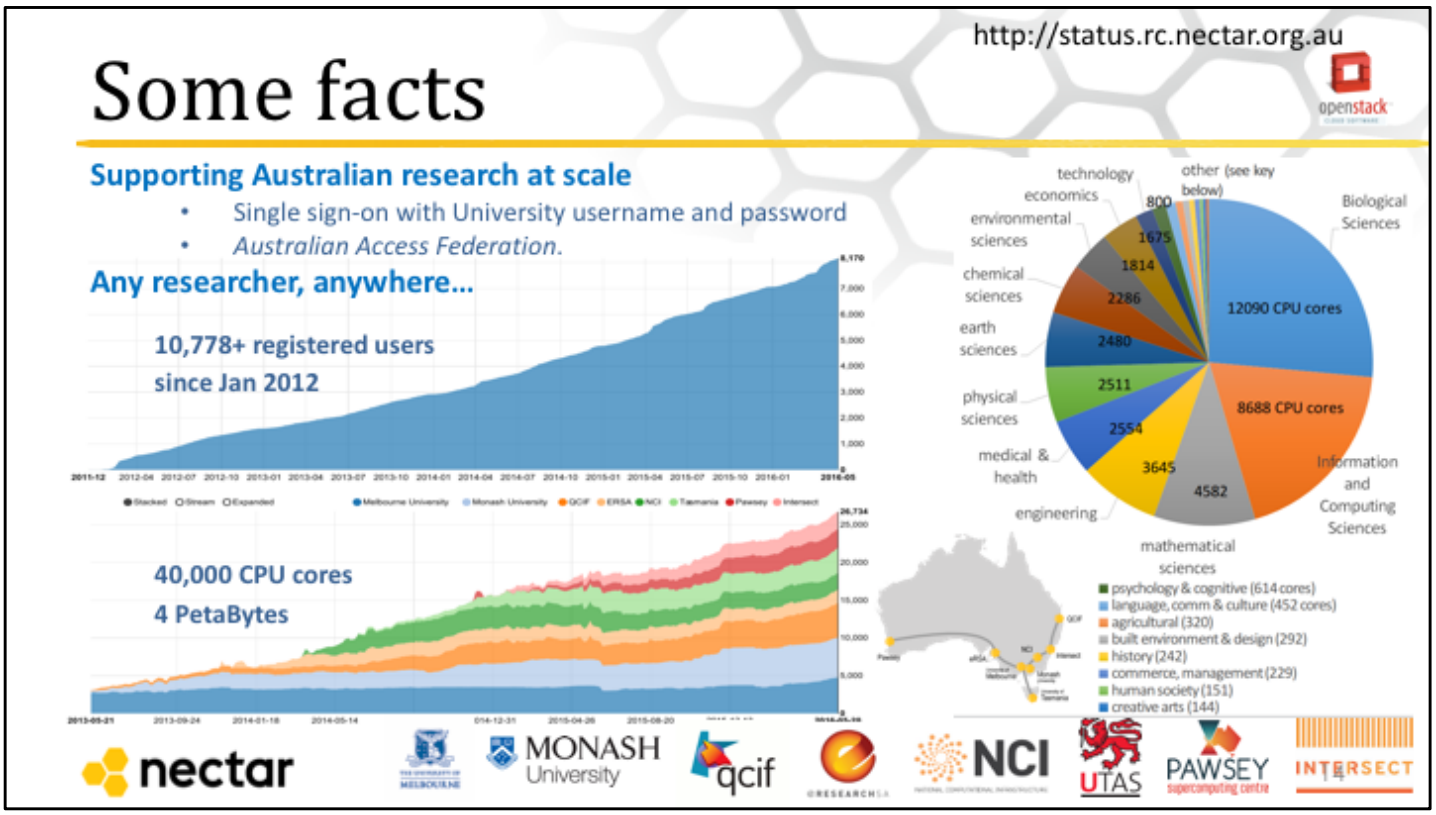
OpenStack Higher Level Services

- Data & analytics – *Trove, Sahara, Gnocchi*
- Application Services – *Heat, Murano*
- Software Defined Networking – *Neutron*
- Storage, backup & recovery – *Swing, Cinder and Manilla*

nectar

13

…



Steady increase

Diversity of use cases

I'd like to spend a bit of time talking about how the national cloud is operated, the structures, experts and community which which support this infrastructure. From what it was 5 years ago, it is now quite a highly evolved, collaborative and streamlined process.

Nectar training Australias openstack expertise.

Nationally ~ 16 FTE spread over 60+ people

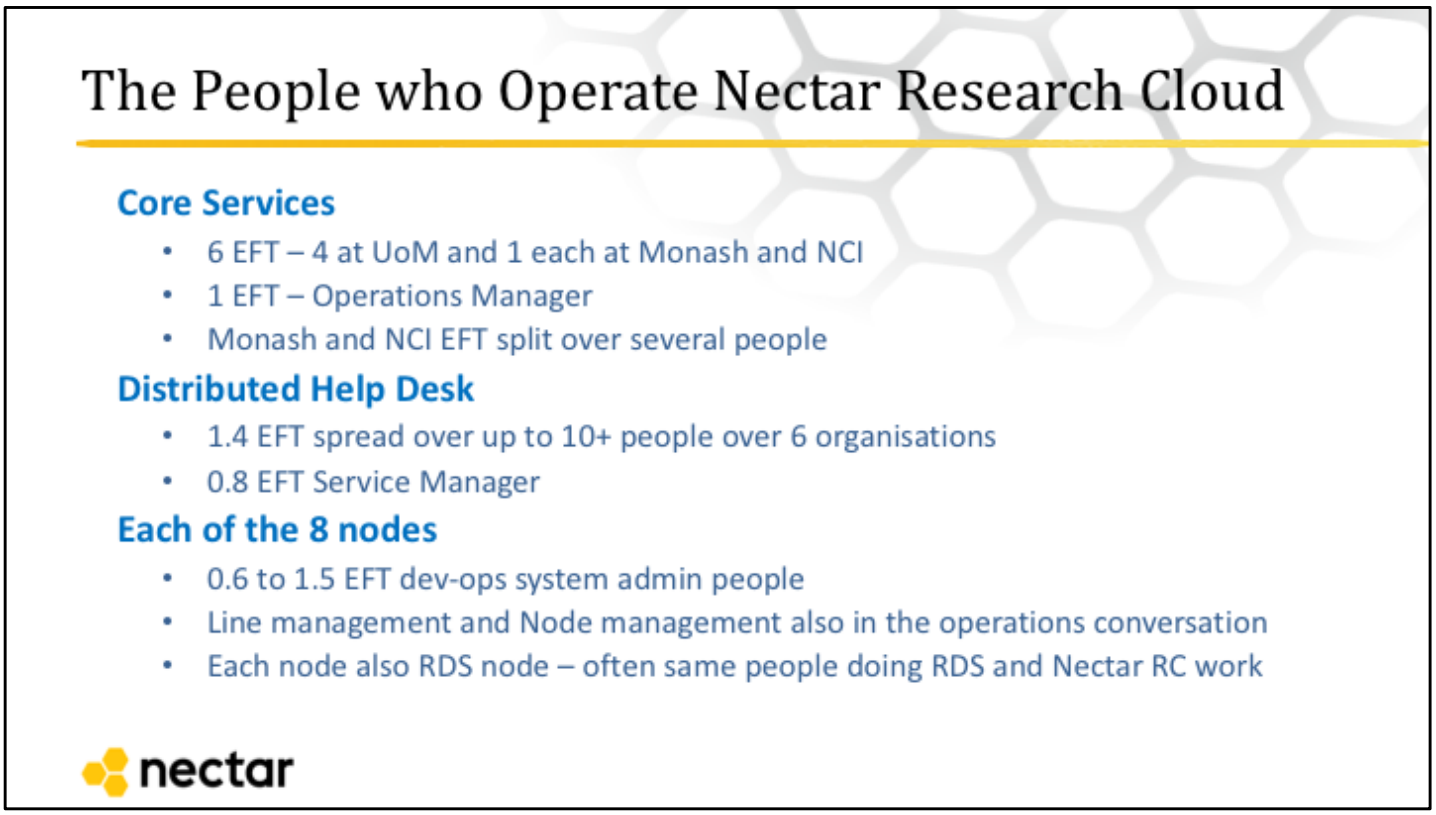
- Core Services
- Distributed Help Desk
- Node operations and Node management
- Nectar Directorate – federation business mgmt

Supporting – 10,000+ registered Nectar Cloud allocations

- plus 1,000's of users within many these allocations

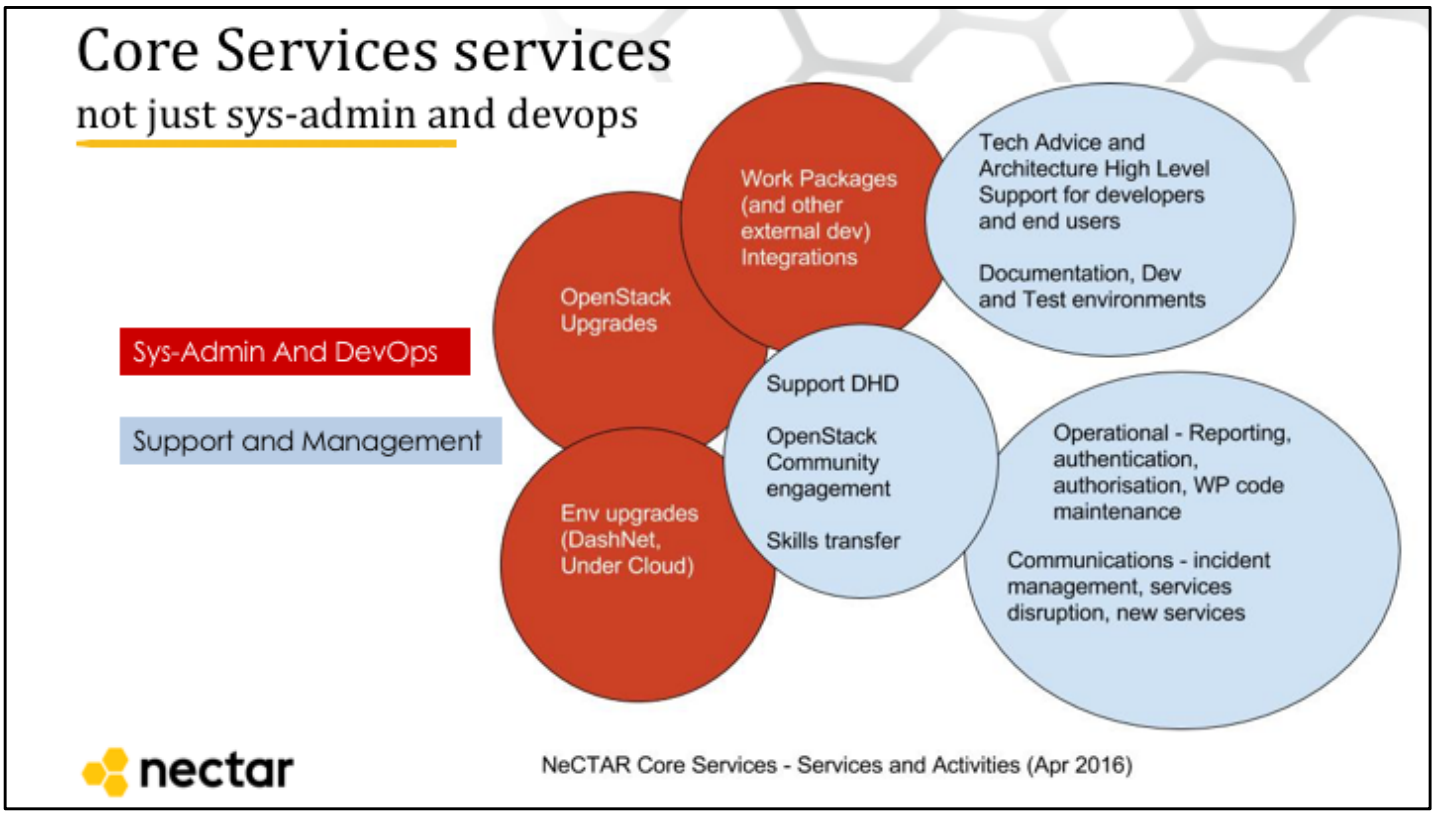
Distributed approach provides for collaboration and skills development and transfer across the nation

Argue that inefficiencies and extra work that come from a distributed approach is more than compensated for by benefits of collaboration and skills development.



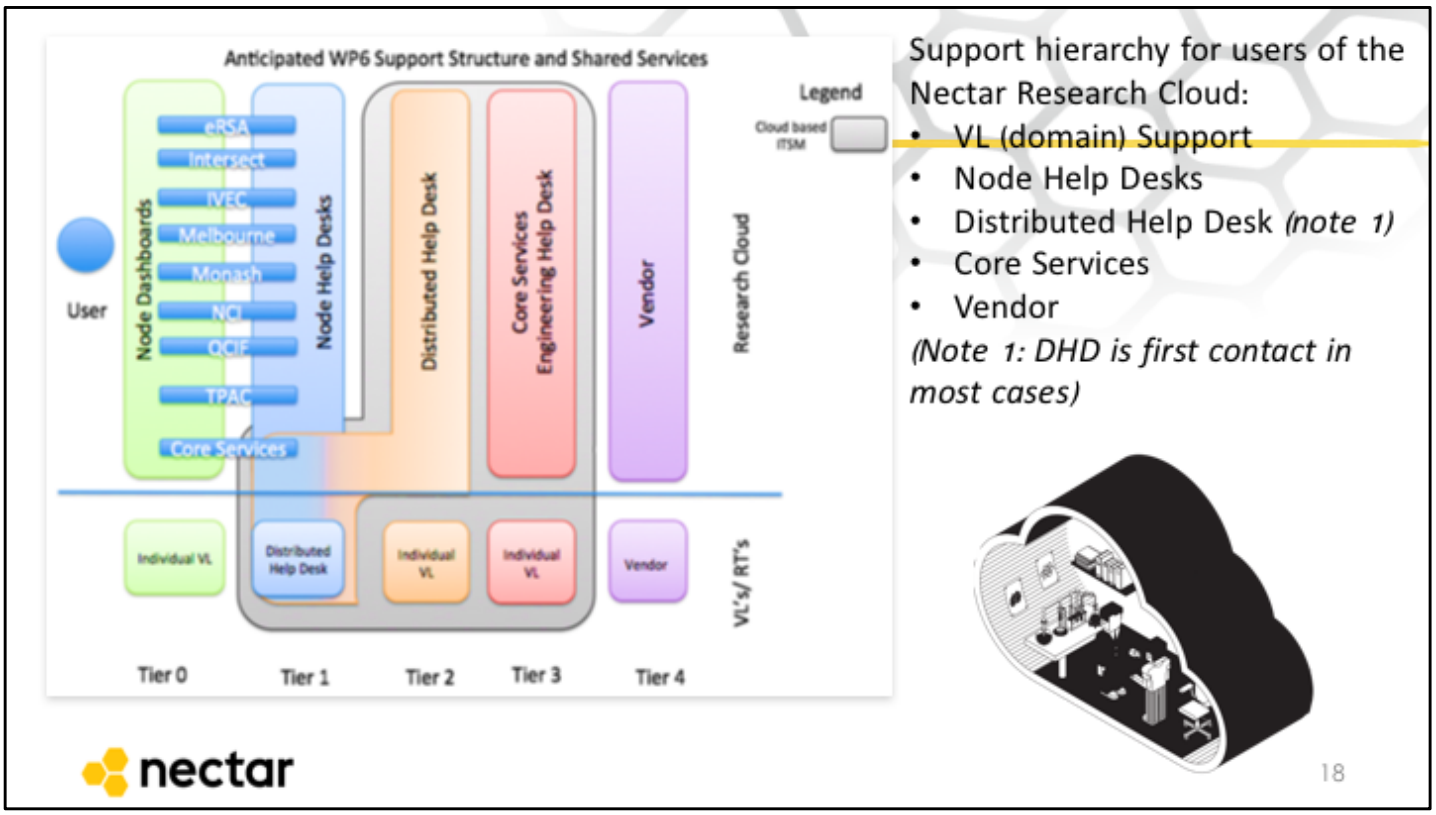
The Monash and NCI Core Services EFT being spread over multiple people does experience service level difficulties from time to time

The DHD is underpinned by a rotating team who take on helpdesk support for the national community …



Core services and the broader national team actually engage in a range of more engineering and technical roles, as well as ‘softer’ areas of activity.

Many people were not aware of the softer areas of activity that Core Services was engaged in. The point was to communicate to Management at Nodes and other management outside of Nectar the bredth of activities that Core Services is engaged in.



We didn't know this on day 1

Activities used to “glue” distributed ops

Six Monthly Tech and Operations Workshops

- Physically bringing Core Services, Node Operators and DHD together for 2 days

Fortnightly RC-ops meetings

- Video Conference of Core Services, Node Operators and some DHD

Weekly Core Services meetings

- Video Conference of Core Services staff from UoM, Monash and NCI

Fortnightly DHD meetings

- Video Conference of 5 nodes participating in the DHD

Monthly Control Group meetings

- Video Conference - Node Management, Nectar Directorate, DHD and Ops Mgmt

nectar

Tools used to “glue” distributed ops

SLACK – Channels for RC-ops, Core Service, DHD, Science Clouds etc.,

ZOOM – Video Conferencing extensively used for meetings and impromptu discussions

Email – of course

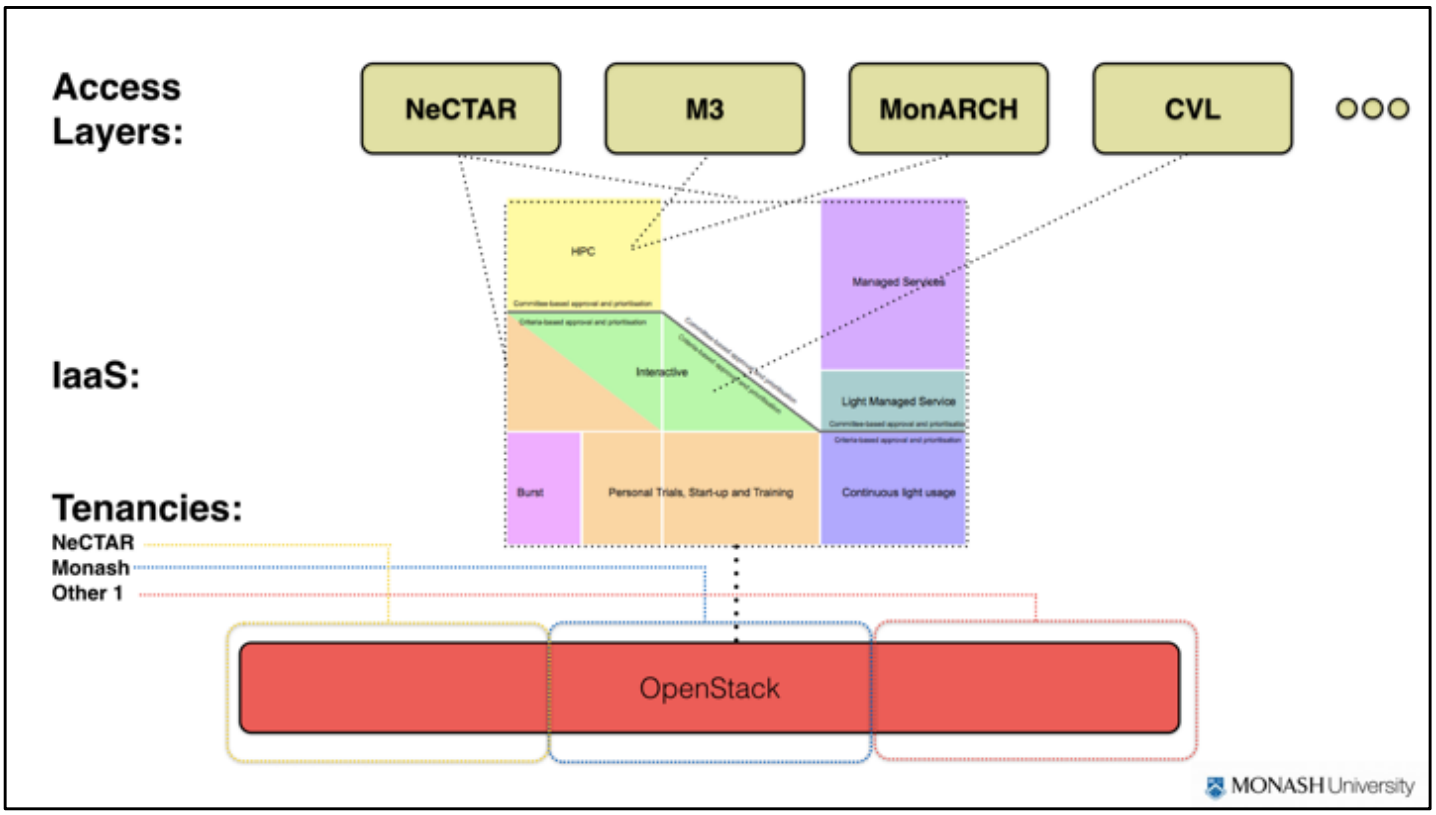
GitHub, Gerrit, Jenkins

Distributed help desk system – FreshDesk based

Documentation

- Wiki for Core Services and Node Operators
- FreshDesk Knowledge Base for end users (and public side of Operations)

nectar



I really powerful thing here is that what this basic underpinning infrastructure allows a proliferation of different focii and business models to exist atop a common platform.\

That is… each of the nodes have evolved independent capabilities atop the common cloud platform.

This is the power here. Common APIs, common operations, common platform, but a diversity of nodes ,each with their own discipline –focussed research interests.

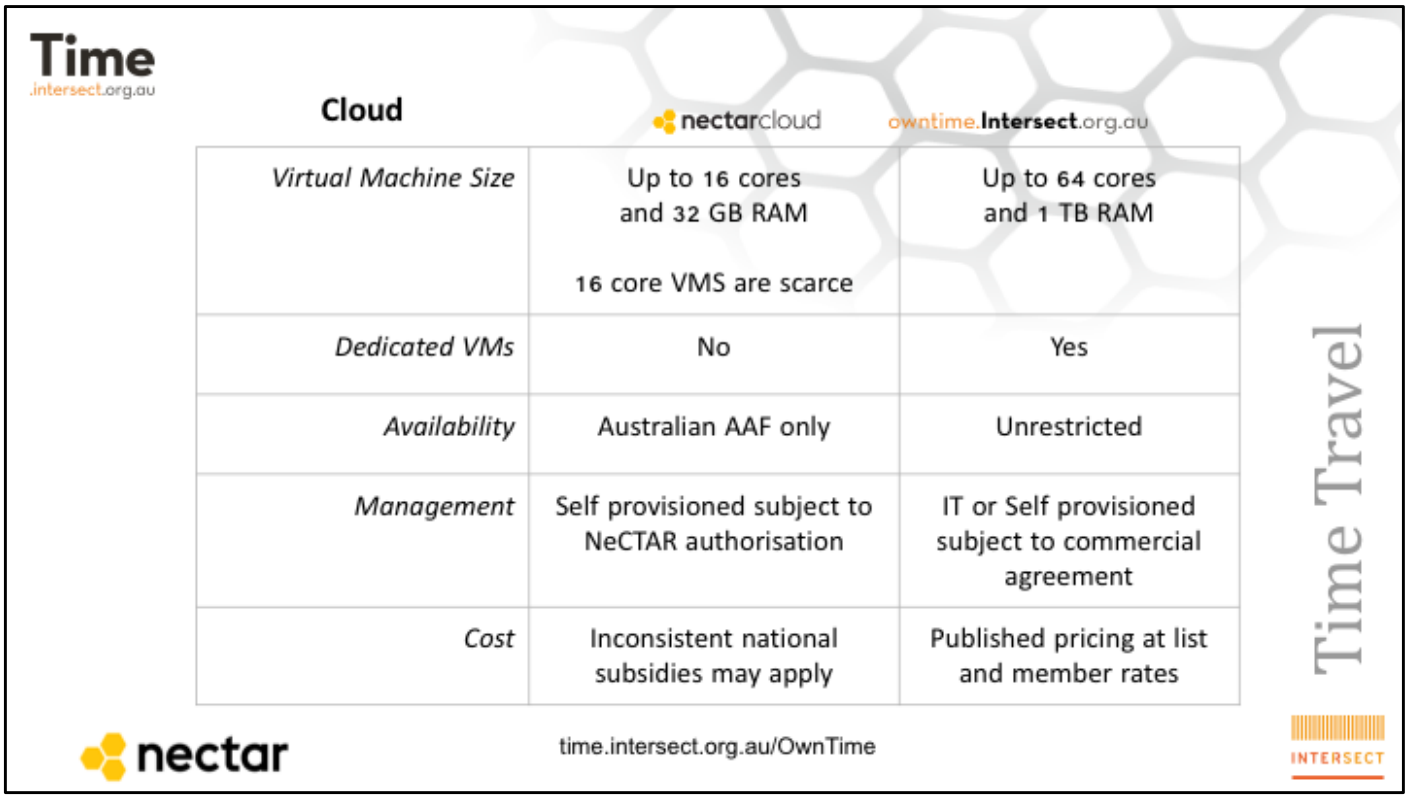
What I want to get across here is that just because you have a common platform which enables and support collaborative research, it doesn't need to look the same everywhere. It can be diverse and representative of local needs.

Monash University

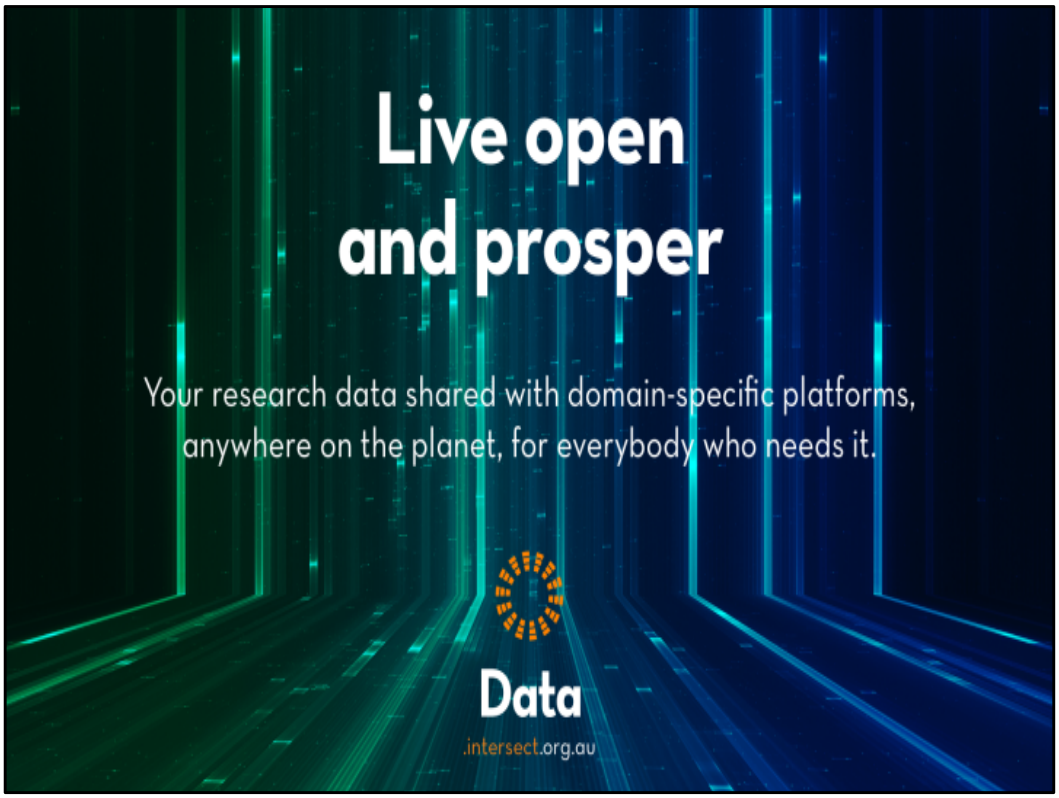
There are always many "front-doors" we need to provide, as sometimes it is the national programs that fund a resource, sometimes it is the university, and sometimes is a research community. These are the "Access Layers". Researchers who use these "front-doors" want the right shape of computing/storage/networking resource for research. Thanks to the NeCTAR

investment, we developed a map of how researchers (under an era of no capacity constraints), of how time-vs-utilisation of a cloud resource relates to flavours. We need, and will always need, more flavours than the commodity (web/container/etc) markets will provide at competitive prices. This is the IaaS layer in the diagram. The entire set of resources of flavours is orchestrated through OpenStack (and in our case Cumulus and Ceph) independent of who funded that resource (tenant)."

Another case in point here is the k



Intersect is an incorporated company, which provides services to the NSW research and govt communities. So the model it utilises is very different to say melbourne or monash. Again, common APIs, common platform, but a different delivery model, different customers, diverse services, etc.

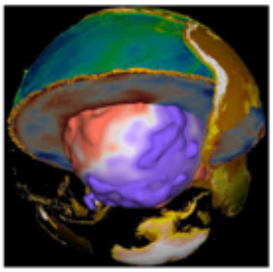


Have gone to great effort, much more than other to market and promote their services to the NSW research and government community in pretty funky ways, through promoting HPC, cloud and data services as time, owntime and data respectively. Userbase is diverse.

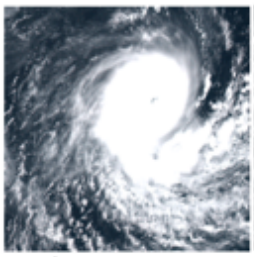
"Whole Earth System" is the particular priority area for research at NCI

Requires cross-domain research

- Modelling Extreme & High Impact events – BoM
- NWP, Climate Coupled Systems & Data Assimilation – BoM, CSIRO, Research Collaborations
- Hazards - Geoscience Australia, BoM, States
- Geophysics, Potential Fields, Seismic, Electromag – Geoscience Australia, Universities
- Monitoring the Environment & Ocean – ANU, BoM, CSIRO, GA, Research, Fed/State
- Agriculture/food security issues

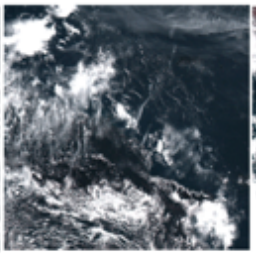


Tropical Cyclones



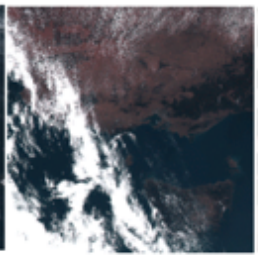
Cyclone Winston
20-21 Feb, 2016

Volcanic Ash



Manam Eruption
31 July, 2015

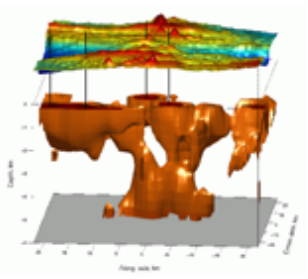
Bush Fires



Wye Valley & Lorne Fires
25-31 Dec, 2015

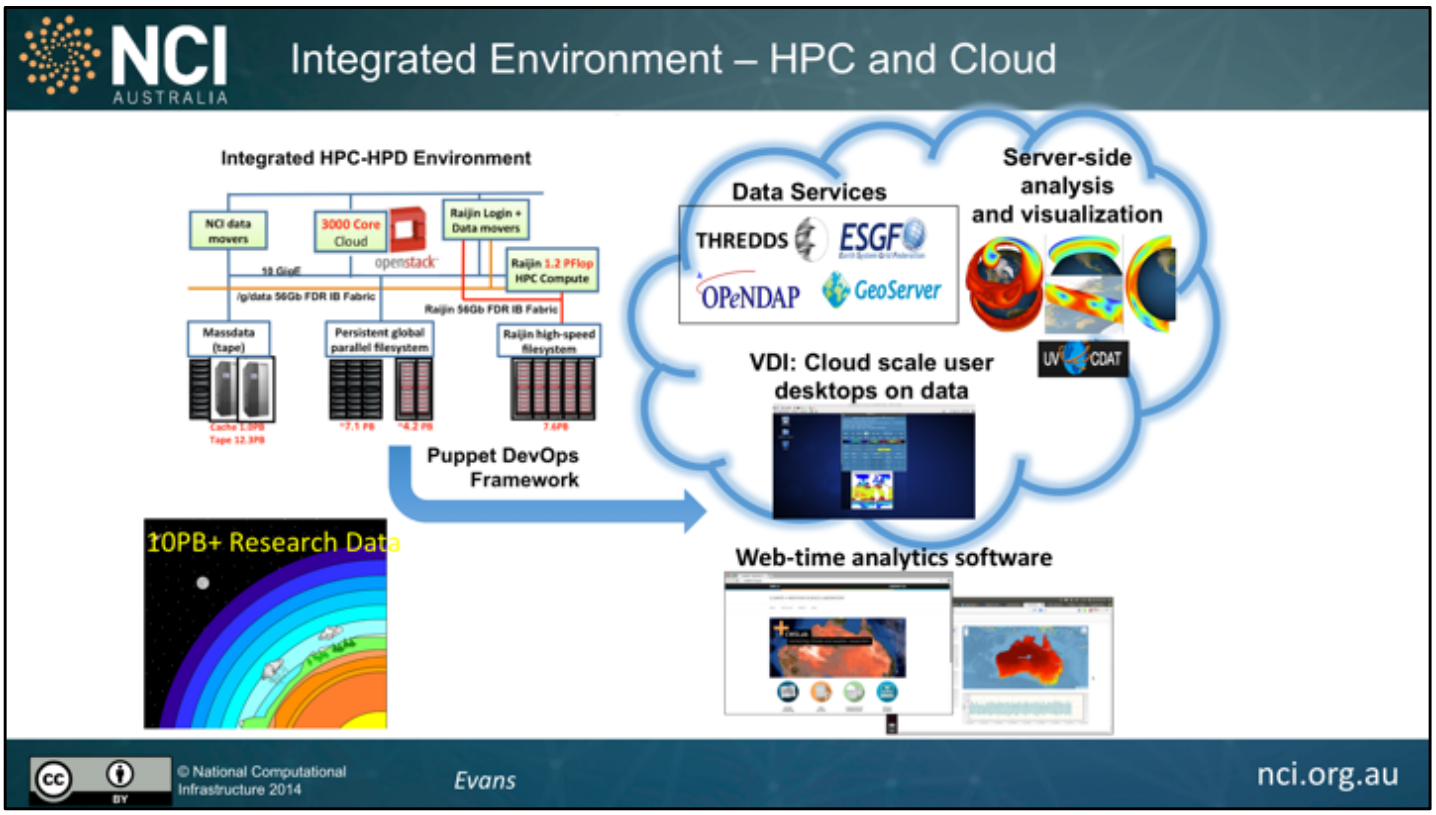
Flooding

St George, QLD
February, 2011



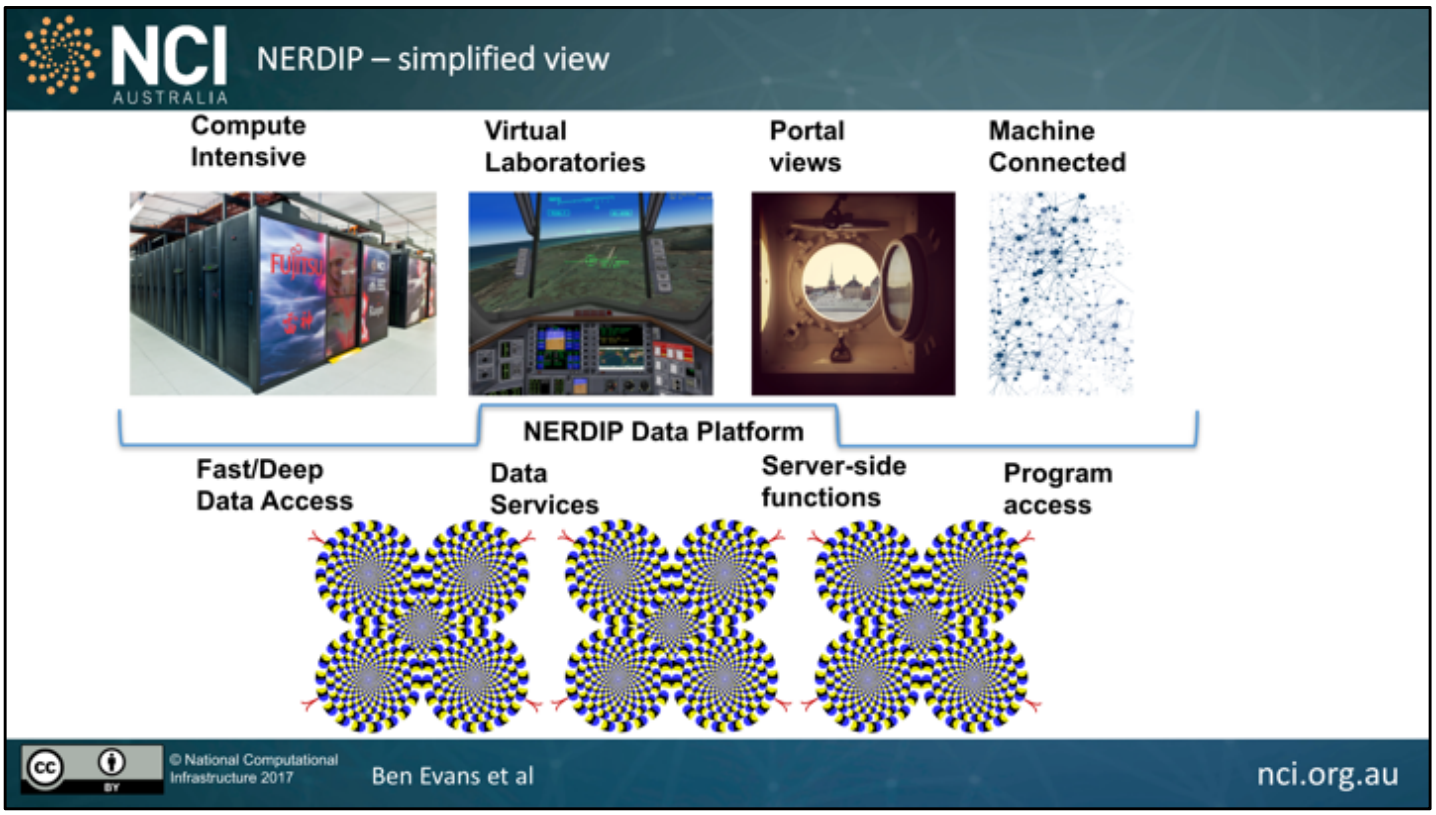
3D Geophysical Models

© National Computational Infrastructure 2017

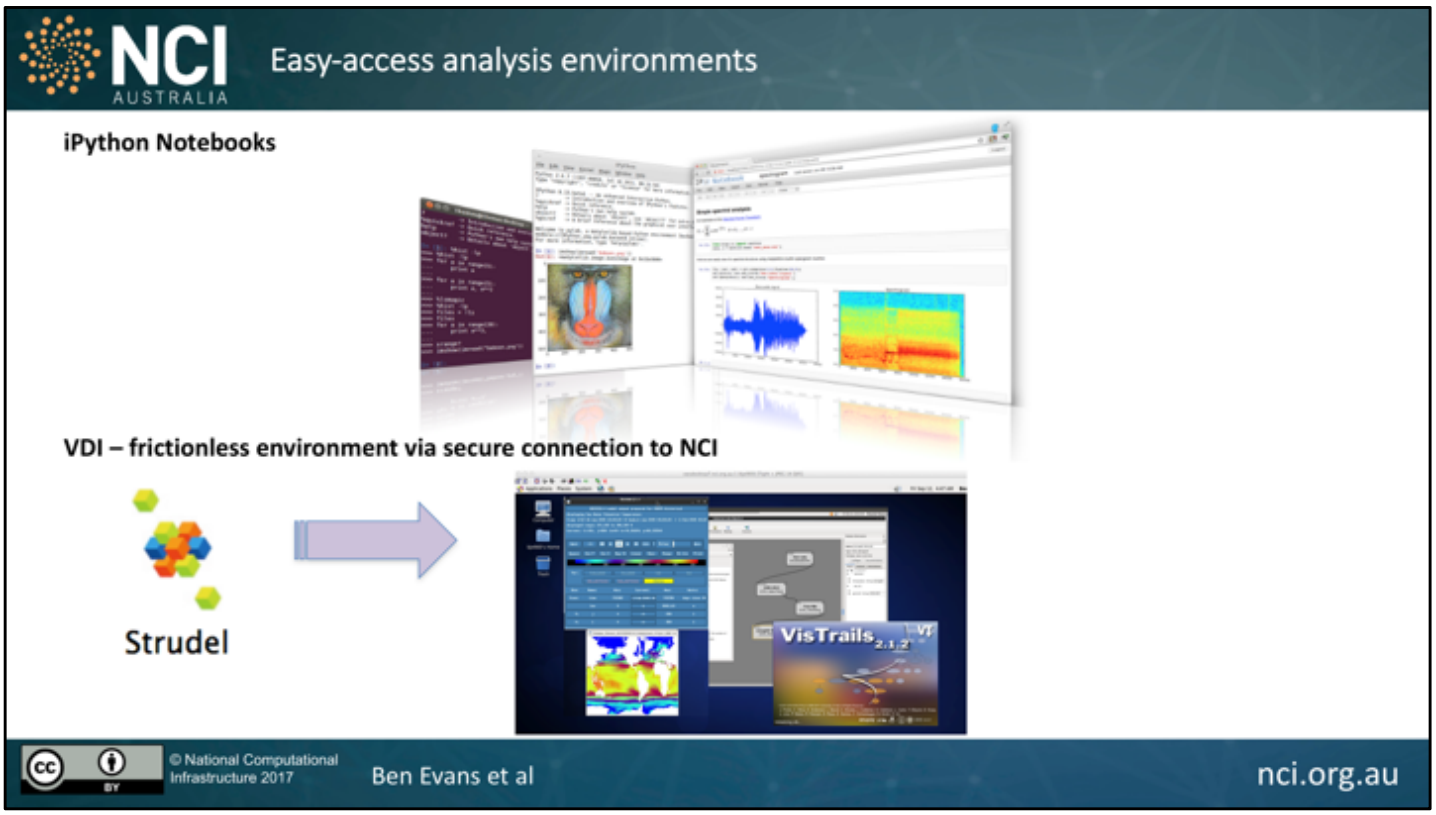


1. NCI has integrated the cloud infrastructure so it plays an integral part in the whole environment. Until Nectar came along, NCI were not engaged in cloud at all.
2. We have built our own Puppet Application level environment which support their primary services
3. We separate out the core application service environment from the non-managed environment. The core application environment has tighter security controls and access to the data e.g. 10+Pbytes of reference data. Most data is made available to non-managed nodes via community standard “data services” rather than “NFS”.
4. There is an ecosystem of core environments – our Virtual Desktop Infrastructure (VDI), and other core services and Virtual labs.
5. There are then various virtual labs and portals that can then be built of our core services

6. … all supported and underpinned by the cloud platform

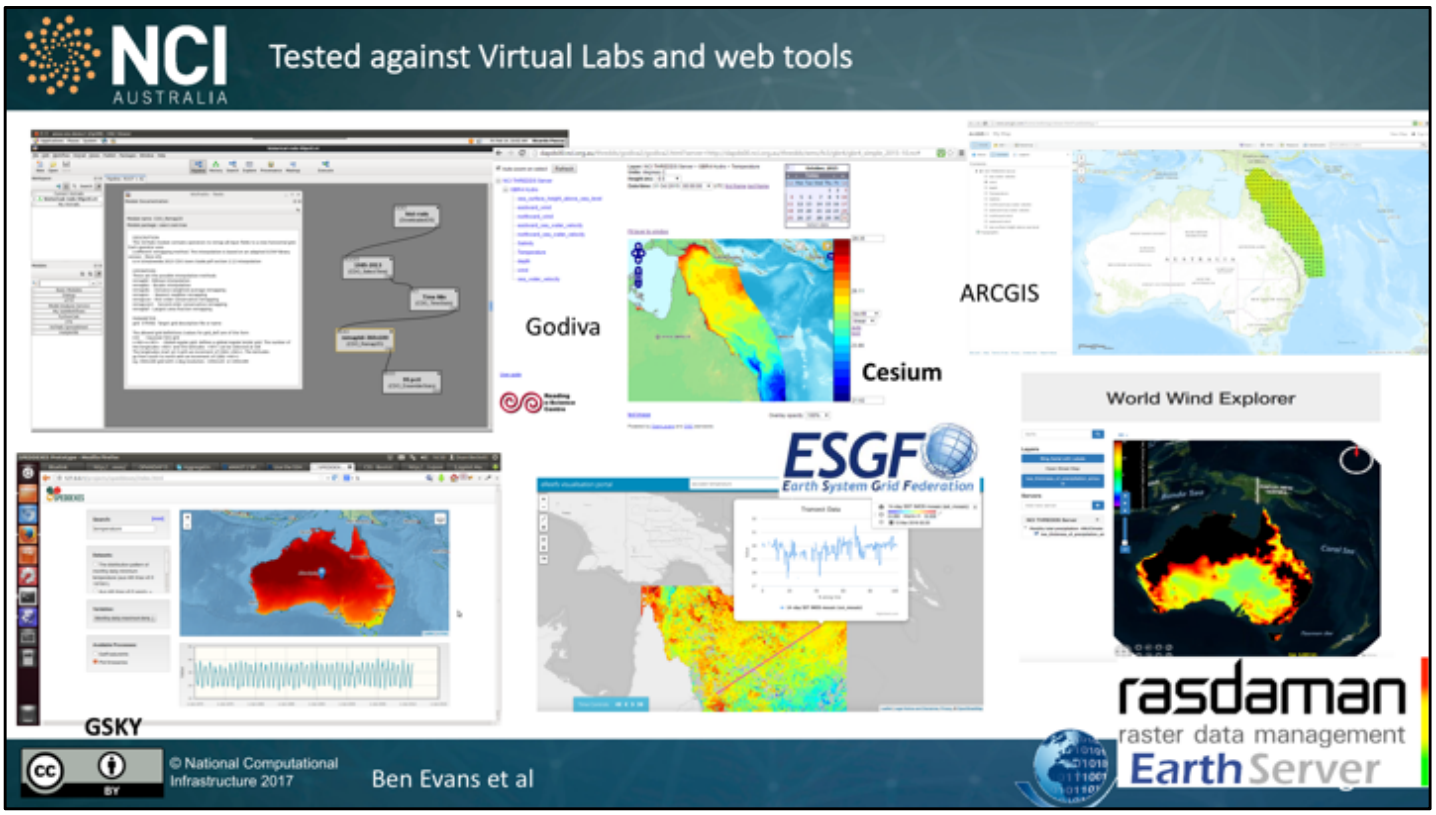


Another, more clear example of our data platform for Earth Systems. It's the National Earthsystems Research Data Interoperability Platform (NERDIP).
The view is that we should provide a foundation layer than works functionally and performantly correct for various types of access. HPC, VLs, Portals and programmatic access via machine connected services deep into the data.



The use of python/jupyter notebooks is now very ubiquitous. We provide a full-features VDI environment that we use, and we make sure that the python notebooks work internally as well as via services.

Our VDI has lots of software tools and libraries well which have direct access to the large repositories of data.



Here is just a small selection of all the range of portals, VLs and workflow environment that are built from this ecosystem.

Cancer Therapeutics CRC

Access to cancer research data, tools and visualisation on the NeCTAR Cloud

Providing access to analysis and visualisation tools, and over 30TB of cancer research data **on the Research Cloud.**
The Nectar choice was easy, and the migration process seamless.

"The service, support and responsiveness that we have received from the Nectar team has been first class, and feels like an extension to our own internal support services."
Paul Reeve, Director of Operations, Cancer Therapeutics CRC.

Plant Energy Biology Centre of Excellence

Building collaboration on the Research Cloud.

Researchers study how plants capture energy from sunlight and how they use that energy to grow and develop.
Hosting collaborations with the Max Planck Institute and the Beijing Genomics Institute – ***on the NeCTAR Research Cloud.***

"NeCTAR makes it much easier, much faster. It means more collaborations — projects that would have just been too hard to go ahead."
Professor Ian Small, Laureate Fellow, West Australian Scientist of the Year 2015.

NeCTAR Research Cloud

*Supporting innovation and collaboration in the **business** of research.*

nectar

Stemformatics

Stem Cell data visualisation on the Cloud.

Find and visualise interesting genes in datasets from leading stem cell laboratories on the **Research Cloud.**

- Over **400 users** nationally
- **100 cores**, multi-site
- NCRIS supported.

29

So they were the node-specific examples, but of course the cloud now underpinns a tonne of research activity around Australia, really supporting innovation in the business of research.

CRC – migrated to nectar cloud while completing potentially Australia's largest research commercialisation agreement. Seamless migration.
Leverage their own IT team – with support from the nectar team.

Plant energy biology – ease of sharing access to data and services.

Stemformatics – agile and redundant infrastructure to support access to key data and services for stem cell research in australia.
Also – international users.

Access to sector-based cloud infrastructure is changing the way research institutions, centres and institutes offer access to data and services.
More agility – more innovation – less risk.

NeCTAR Virtual Laboratory Case Studies

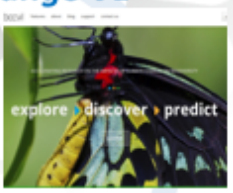
Genomics VL

"This is the best exemplar of this kind of platform in the world… Genomics capability for the masses."
Associate Professor Andrew Lonie, Director, EMBL-ABR.

The **Peter MacCallum Cancer Centre** is using the GVL in the NeCTAR ***Research Cloud***, providing instant access to Genomics tools and data for Australian biologists

Biodiversity and Climate Change VL

"..decreases the time to complete biodiversity analysis from 2 months to 5 minutes, supporting new applications in research, government and industry."
Professor Brendan Mackay, Director, Griffith Climate Change Response Program



Accelerating biodiversity–climate change modelling across large disparate datasets quickly and easily on the ***Research Cloud***.

Virtual Laboratories are:

- *Accelerating* research
- Leveraging the **Research Cloud** for wide access and **collaboration**
- Bringing together **observation** and **modelling**

nectar

Marine VL

"MARVL enables researchers to start thinking about their problem sooner."
Dr Roger Proctor, Director e-Marine Information Infrastructure Facility.

Ocean observations and modelling for marine and coastal environments
Ian Coghlan is studying coastal erosion. **MARVL saves him 3 months effort** to access local data, wave model simulations and computing resources.

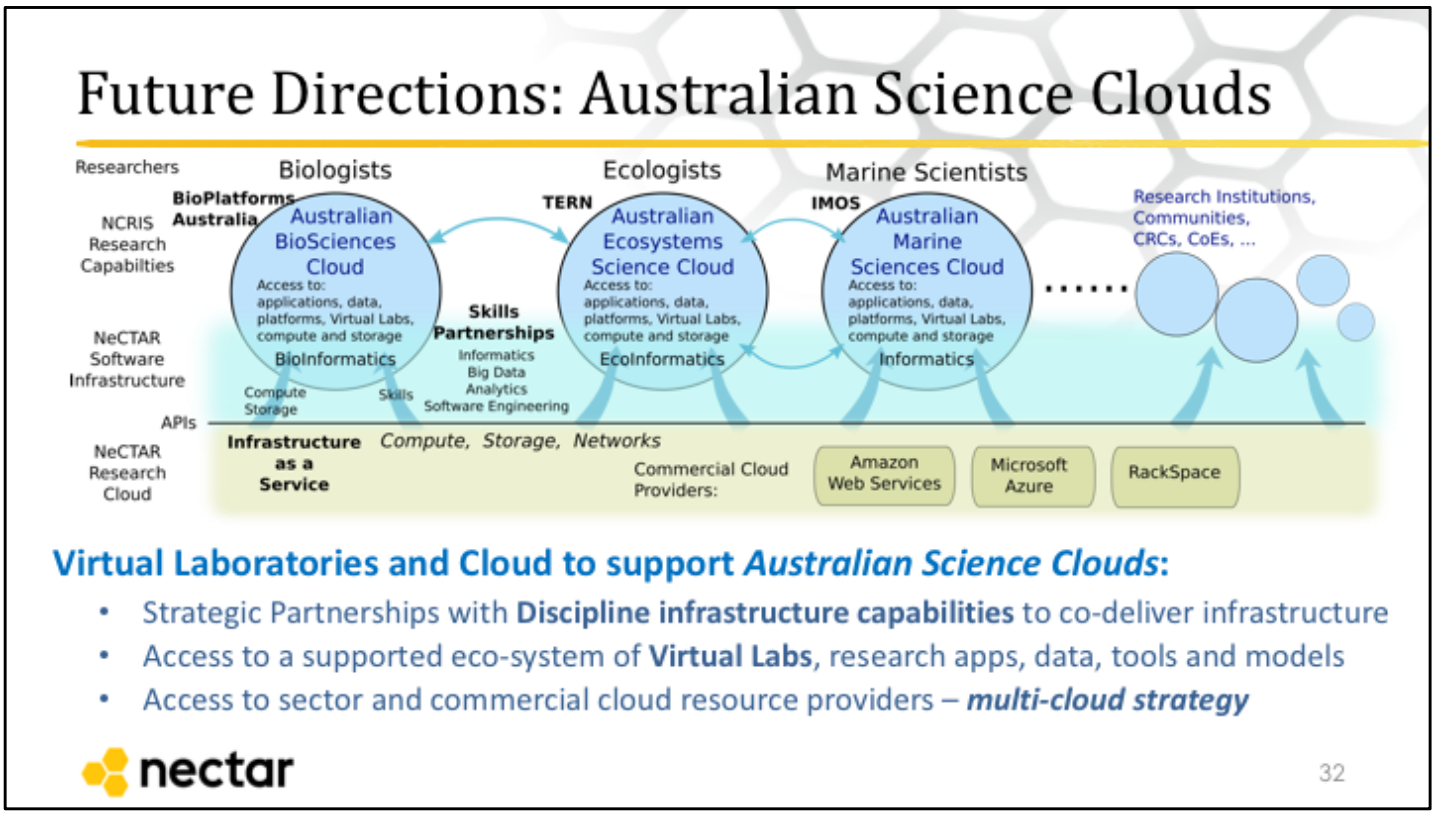
30

Future Directions: International Science Clouds

NeCTAR pioneered the way – but others have followed:

- Over 22 International Science Clouds established (OpenStack-based)
  - Jetstream at Indiana U is 15000 cores
  - Chameleon NSF cloud is 16000 cores
  - UTSA Open Cloud Inst. is 12000 cores
  - Grid.5000 in France is 8000 cores
  - CERN cloud is 151,000 cores, 60,000 more in 2016
  - CLIMB UK cloud is 7680 cores (4 Unis)
  - Compute Canada Research Cloud (distributed national resource)
- NeCTAR is pursuing ***interoperability*** with International Science Clouds:
  - OpenStack Scientific Working group
  - Global e-Infrastructures Interoperability Working Group
  - Boston Declaration (next week @ the OpenStack summit)

nectar



A strategic direction that was established by the Nectar program about 18 months ago was the move ‘up the stack’ to science clouds.

Nationally, we’re in the process of Establishing 3 Science Clouds, Australian BioSciences Cloud, Australian EcoSystems Science Cloud, Australian Marine Science Cloud.

A science cloud in general terms does three key things:

(1) Leverages the common infrastructure in the Research cloud and the Virtual Laboratories programmes and tunes portions of it to meet the needs of specific research communities.

(2) They are a vehicle for investment. They give us a label to allow a domain specific focus for investment into common infrastructure. This supports strategic partnership with the research community to shape the infrastructure to meet their specific needs.

(3) Science Clouds are a vehicle for organizing cross-institutional resourcing. What the VL program, and these science cloud program have done and will continue to do is join together experts from a diverse range of institutions who want to work to a common discipline cloud.

MARINE CLOUD - To provide marine scientists and students a robust framework of tools and resources with easy access to open data in a standard environment. Designed to reduce the time for tool and code preparation as well as the data discovery process.

ECO CLOUD - A place that all environmental, climate and biodiversity data is **dynamically and natively** available to common analysis tools. A place where researchers can **view, query** data **regardless of location**. A place where researchers can get accurate and up-to date **scientific and technical** user support.

BIOSCIENCES CLOUD - Define a common platform and set of services upon which national significant outputs such as datasets, tools and workflows can be published and utilised by the national biosciences research community.

Make the cloud more attractive for coinvestment

1. There is a real complexity to openstack. It's a diverse set of things, it's not just a software tool you install. Release management, testing, debugging, development, distributed helpdesks, engaing with users, these are all things which must be appreciate

2. The overall cloud function is lean because of this, and its an ongoing function that takes time and effort..

3. i.e. the 2 core argument… and T&C's only implemented now…

4. Node success factors…

5. Diversity of approaches and nodes..

6. Building a core, accessible team of engineering expertise and prowess really helps lift the competency of everyone. It's very easy to fumble around with openstack for months. Having access to an extended network of experts is extremely beneficial. And it builds up expertise over time in a distribute way…

7. Contribute / Go to the summits / Learn / connect

Thank you ;-)

OpenStack workshop tomorrow – Justin & Sam.

Let's keep working together to build open, interoperable and international-scale cloud platforms for research.

@DrStevenManos

nectar