What Data and Compute Resources can I access at NCI

Dr. Ben Evans
Associate Director, REI
National Computational Infrastructure (NCI) Australia

What is NCI?

NCI Australia is the **National Computational Infrastructure**

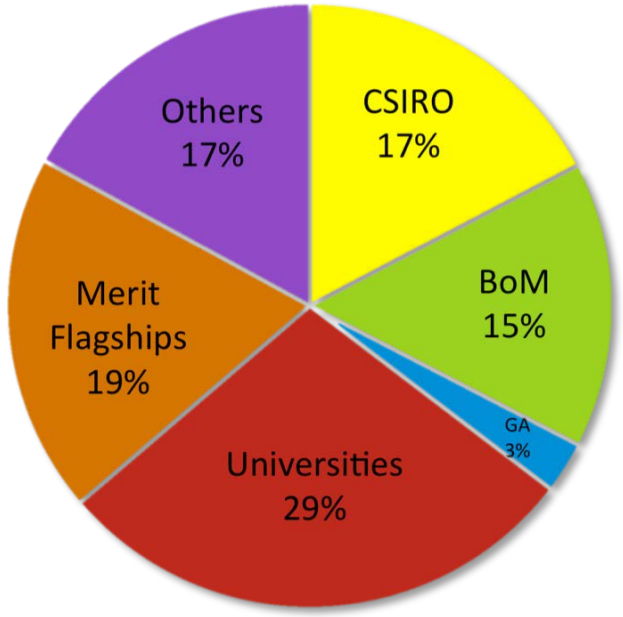
- This infrastructure, and the extensive services built on top of it, are delivered by a team of over 60 staff, know nationally and internationally for their expertise.
- Grew out of ANU Facility (mid 1980s)
- Now is a NCRIS facility – National Research Infrastructure

What is NCI?

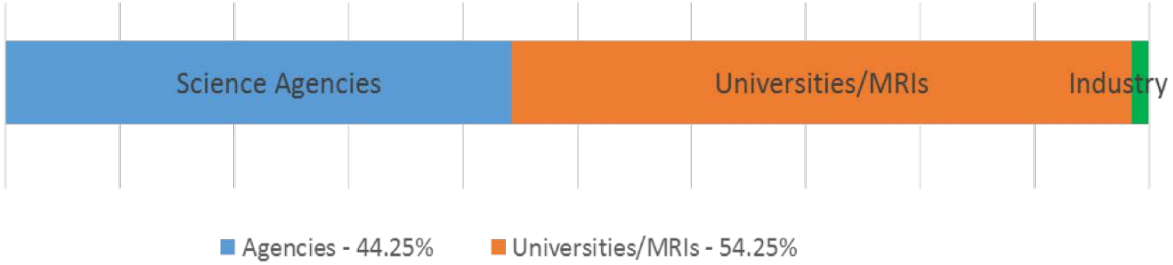
- Partners and S/C – ANU (host institution), Bureau of Meteorology, CSIRO and GA
- Support many partners and stakeholders, including
  - ARC Centres of Excellence (notably, Climate Extremes)
  - NCRIS areas (notably, AuScope)

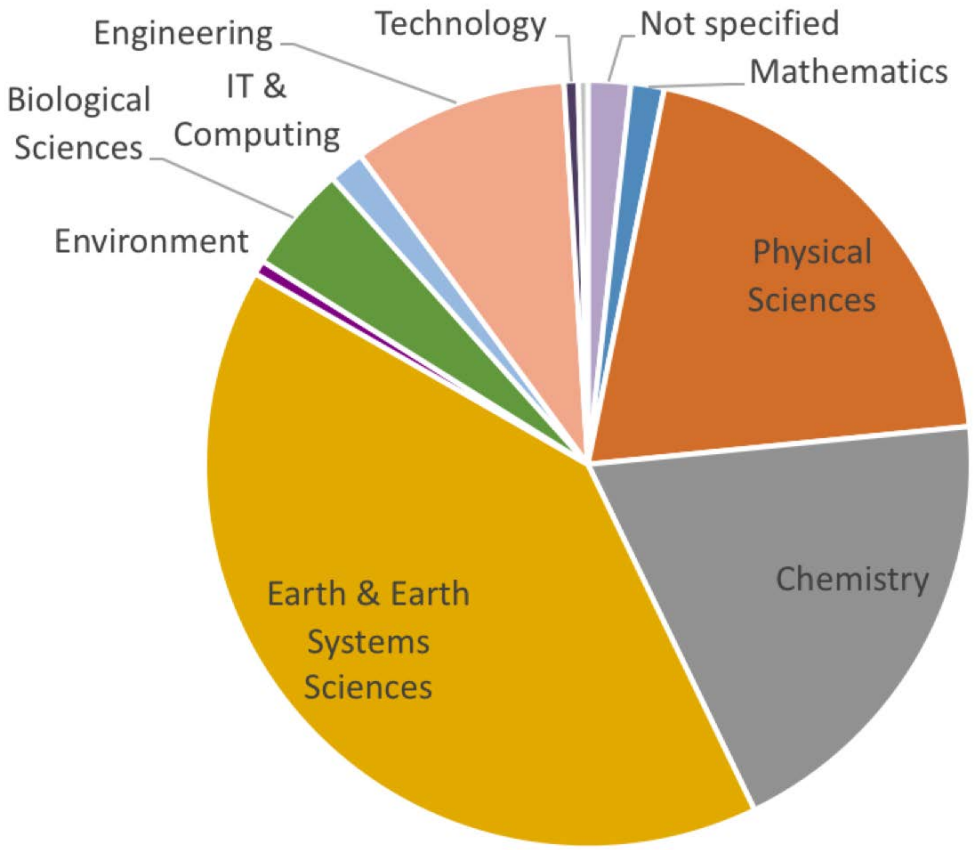
Users of NCI

Distribution by Share Type



Distribution by Organisation Type



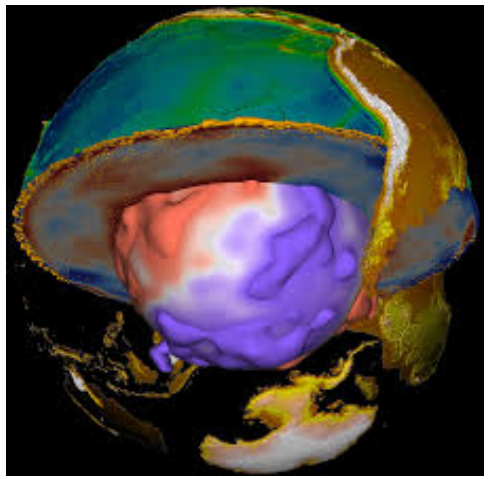


ANU access – primarily through ANU share
https://anumas.nci.org.au/2018/

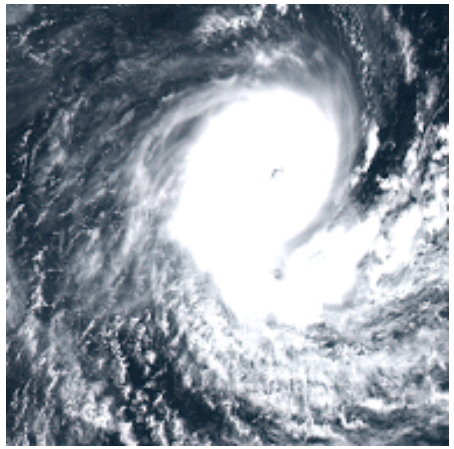
Examples of priority areas for research at NCI

Priority Research areas in Earth Systems

- 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

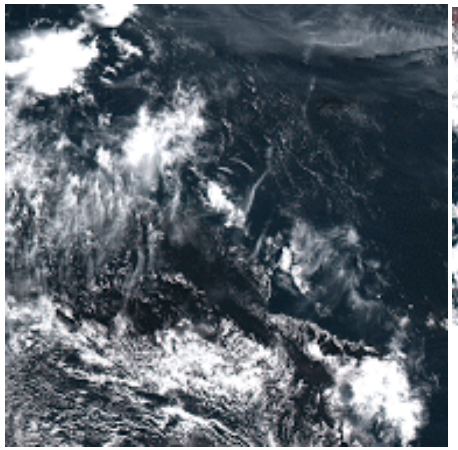


Tropical Cyclones



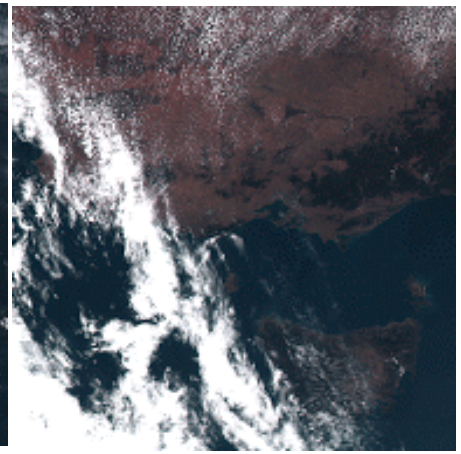
Cyclone Winston
20-21 Feb, 2016

Volcanic Ash



Manam Eruption
31 July, 2015

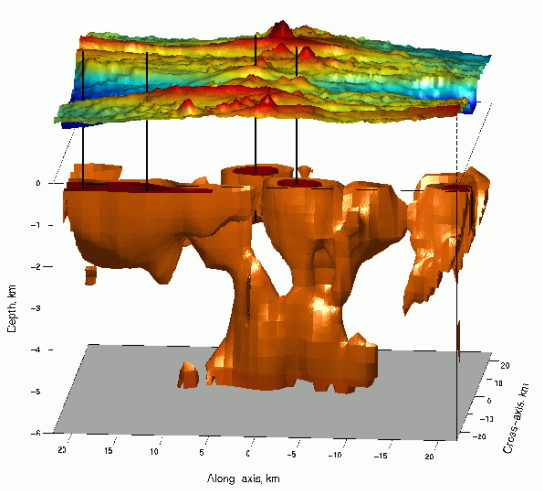
Bush Fires



Wye Valley & Lorne Fires
25-31 Dec, 2015

Flooding

St George, QLD
February, 2011



Supercomputer: Raijin

84,656 cores) in 4416 compute nodes
- Intel Xeon Sandy Bridge 2.6 GHz,
- Broadwell 2.6 GHz

GPUs:
- 120 NVIDIA Tesla K80 GPUs in 30 nodes
- 8 NVIDIA Tesla P100 GPUs in 2 nodes

32 Intel Xeon Phi in 32 compute nodes
- 64 core Knights Landing, 1.3 GHz

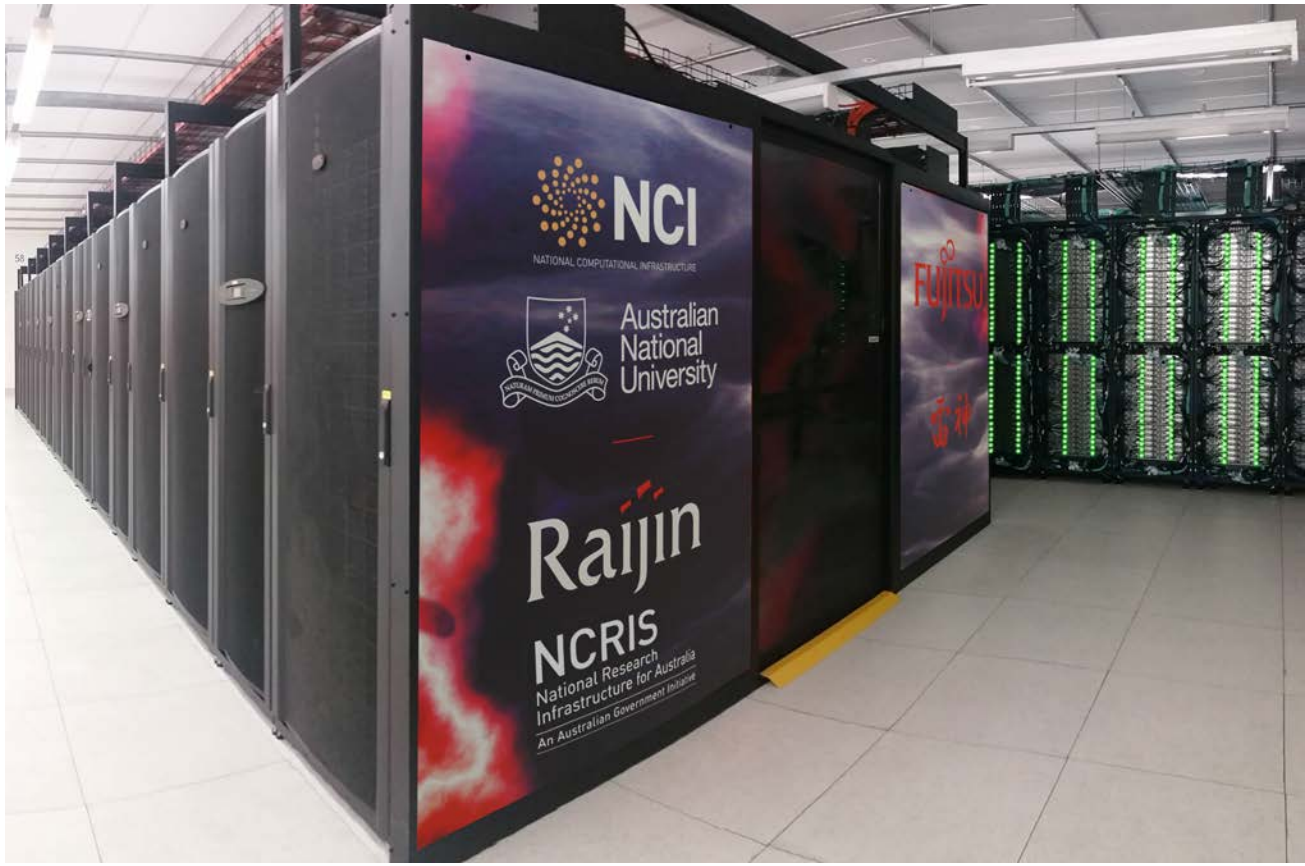
4 IBM POWER8 nodes
- 64 cores running at 4.02GHz)

300 Terabytes of main memory

Full Fat tree interconnect (up to 100 Gb/sec)
- Hybrid FDR/EDR Mellanox Infiniband

~8 Petabytes of scratch storage

PBS batch operating system



NCI AUSTRALIA Raijin

84,656 cores) in 4416 compute nodes
- Intel Xeon Sandy Bridge 2.6 GHz,
- Broadwell 2.6 GHz

GPUs:
- 120 NVIDIA Tesla K80 GPUs in 30 nodes
- 8 NVIDIA Tesla P100 GPUs in 2 nodes

32 Intel Xeon Phi in 32 compute nodes
- 64 core Knights Landing, 1.3 GHz

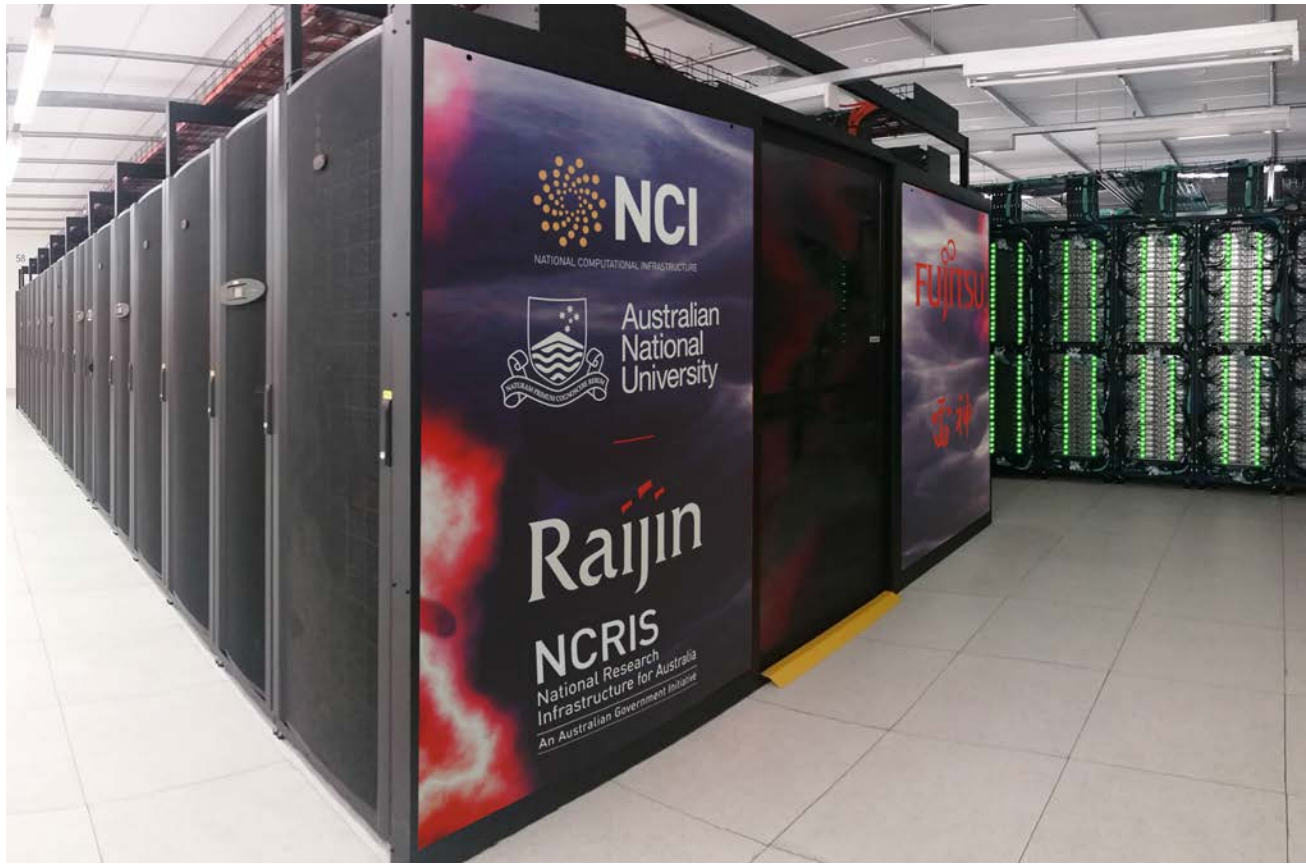
4 IBM POWER8 nodes
- 64 cores running at 4.02GHz)

300 Terabytes of main memory

Full Fat tree interconnect (up to 100 Gb/sec)
- Hybrid FDR/EDR Mellanox Infiniband

~8 Petabytes of scratch storage

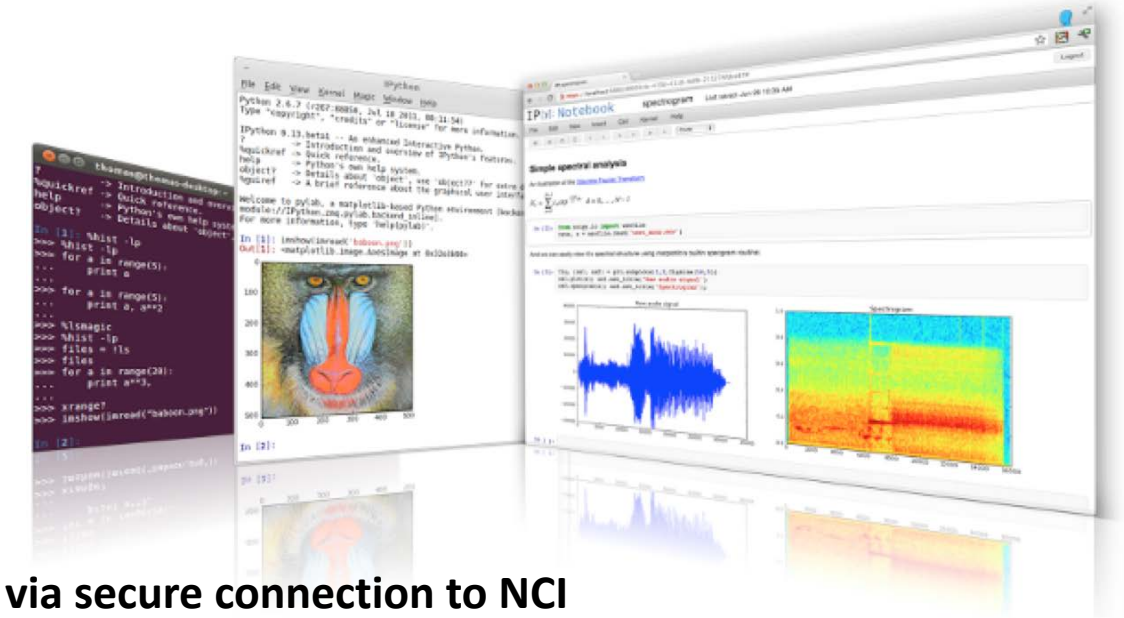
PBS batch operating system



Significant upgrade replacement in 2019

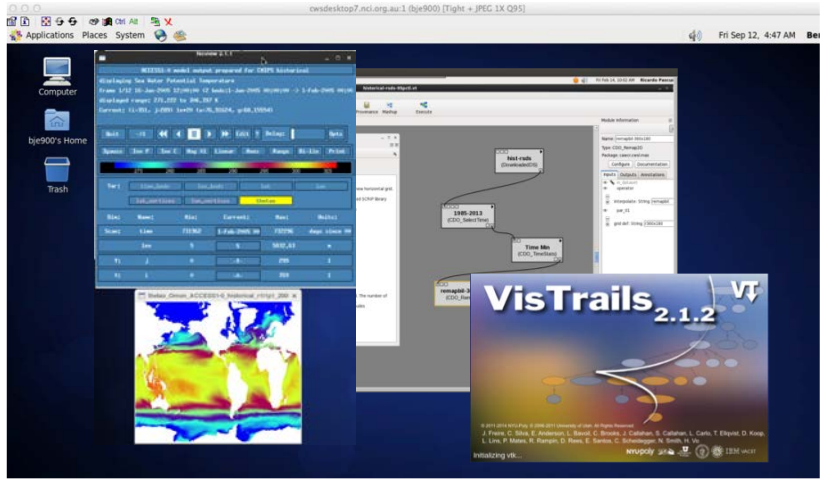
VDI: interactive analysis cluster

iPython Notebooks

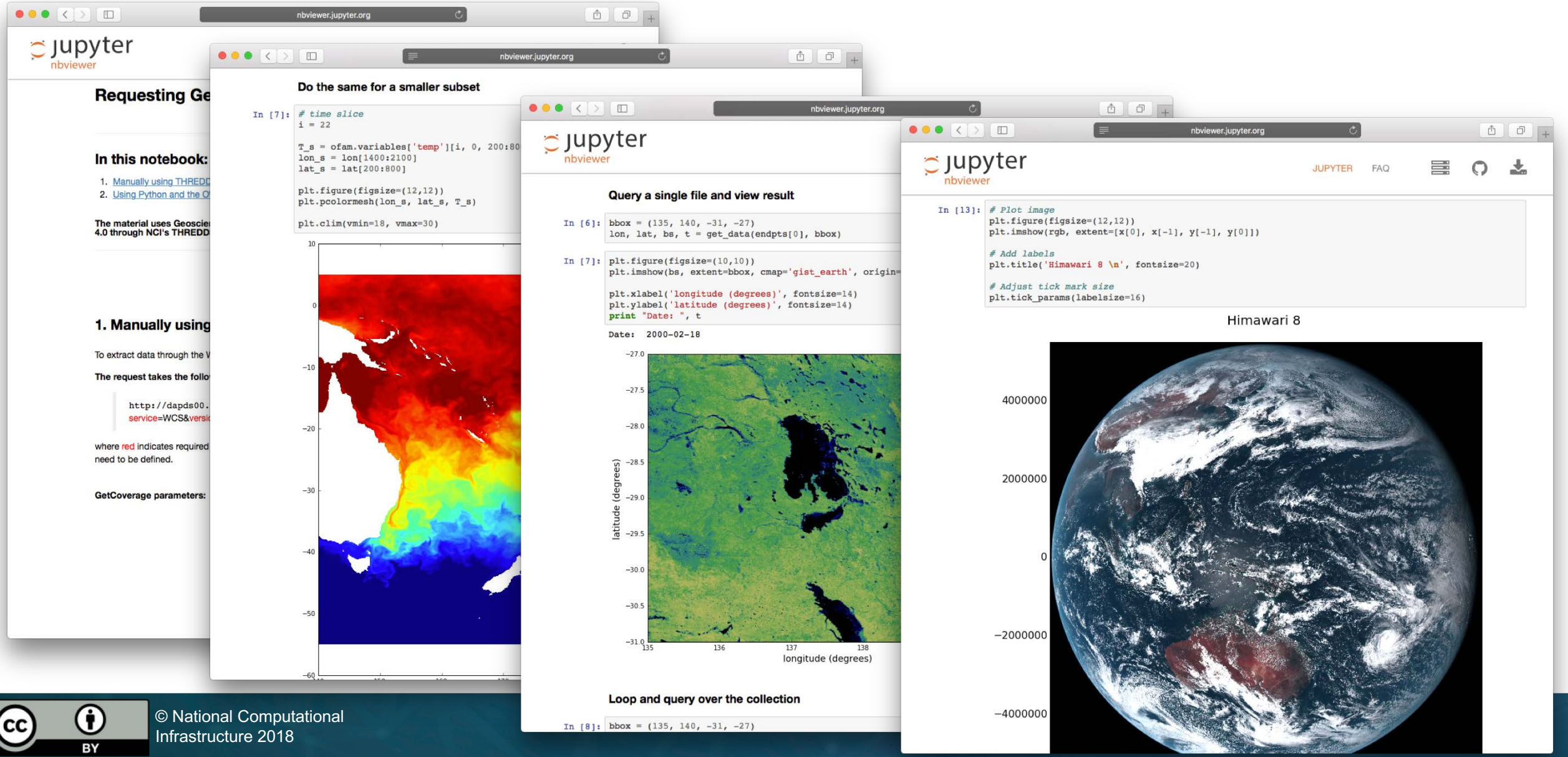


VDI – frictionless environment via secure connection to NCI

Strudel



User reference material



NCI Software

A large collection of maintained software:

- over 1000 packages, libraries and software suites
- General packages: https://opus.nci.org.au/display/Help/Software

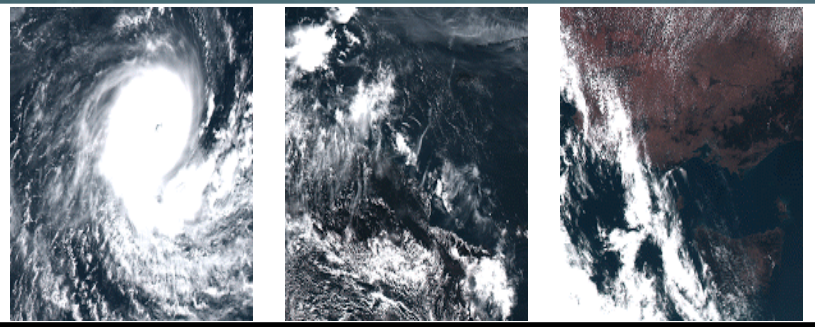
Deep Application activity on code improvement (related to priority areas))

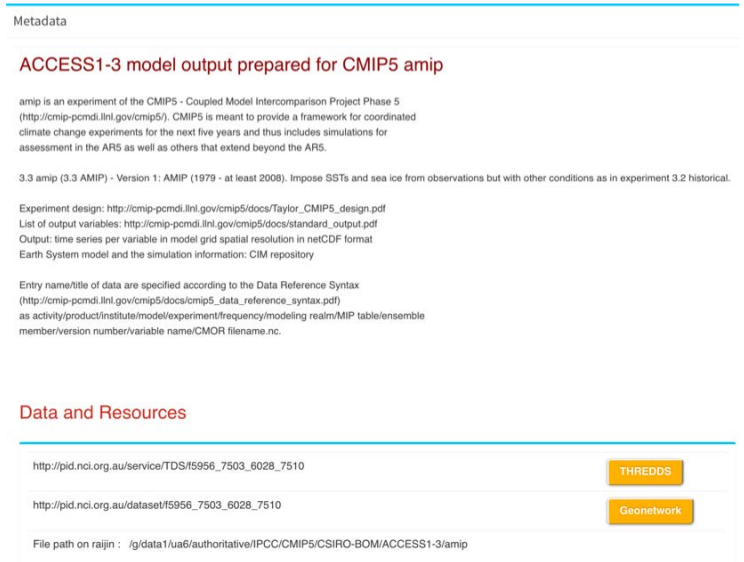
Earth System models			BOM – NCI and Ops CSIRO – NCI CoE/ANU/UNSW – NCI IMOS - NCI
ACCESS APS2 (Global, Regional, City)	ACCESS-CM2	ACCESS-S1	
ACCESS-OM2	UM 10.6/7 N512, N728, N1024, N1280	WRF	
MOM5 0.25, 0.1 (inc. CICE5 and SIS1)	MOM6	NEMO 0.25	
4DVAR N216, N320	EnKF-C	BODAS	
GC2/GC3	ROMS	GITM	
WaveWatch3	NOAA MOST	KAU-GCM	
Geophysics			AuScope ANU/Syd/Melb GA
eScript	Underworld	Badlands/pyBadlands	
JAGURS	OpenQuake	Mod3D (MT)	
Earth Observation			GA – NCI CSIRO – NCI/server-side
AGDC processing pipelines	Landsat	GSKY (in progress)	
Genomics			
GATK			
Astrophysics			ANU - Astro3D
Gadget 3 (prep 4)			

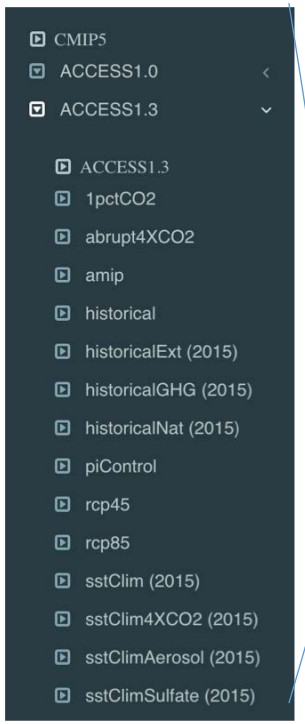
NCI National Earth Systems Research Data Collections

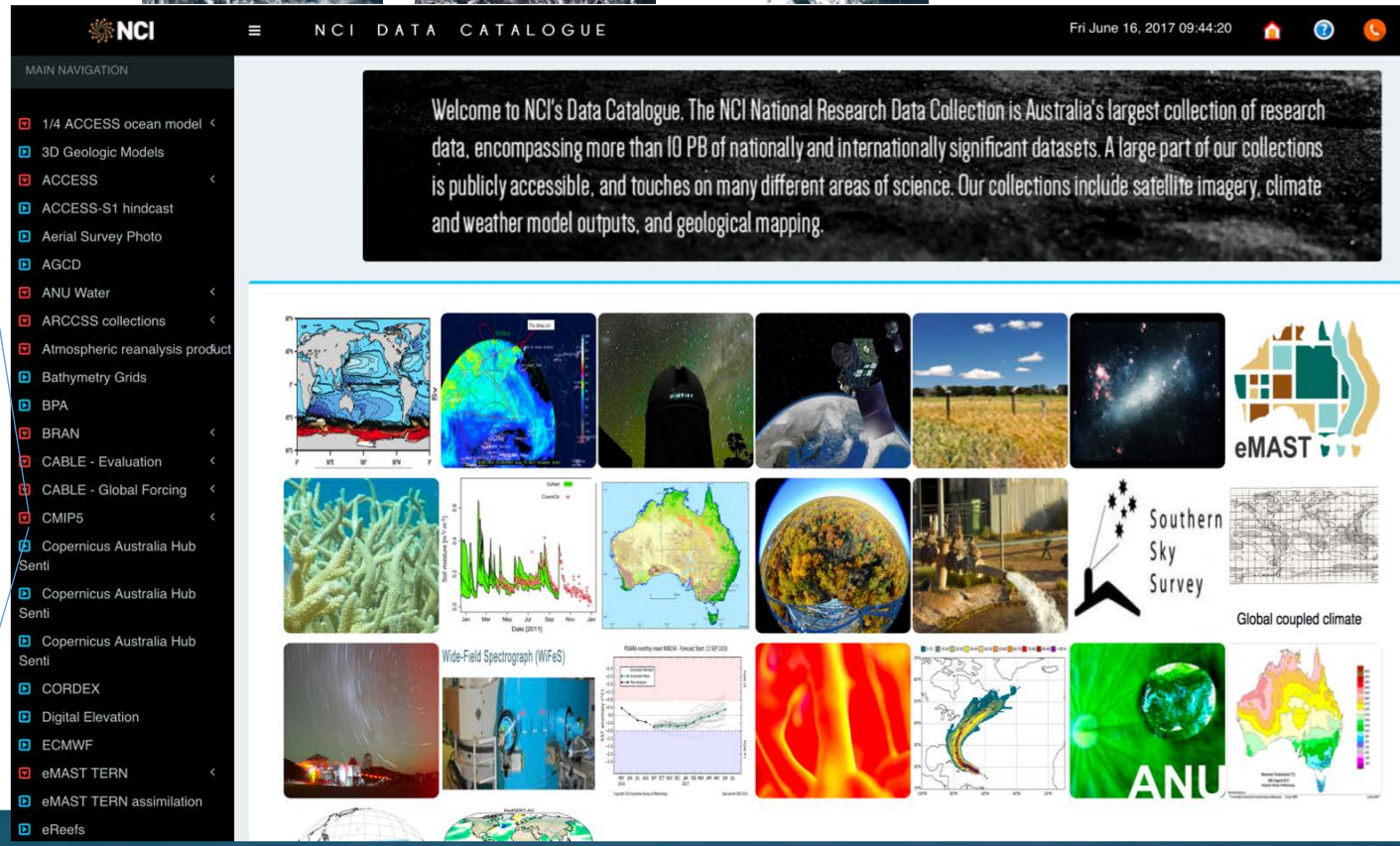
1. Climate/ESS Model Assets and Data Products
2. Earth and Marine Observations and Data Products
3. Geoscience Collections
4. Terrestrial Ecosystems Collections
5. Water Management and Hydrology Collections

https://datacatalogue.nci.org.au

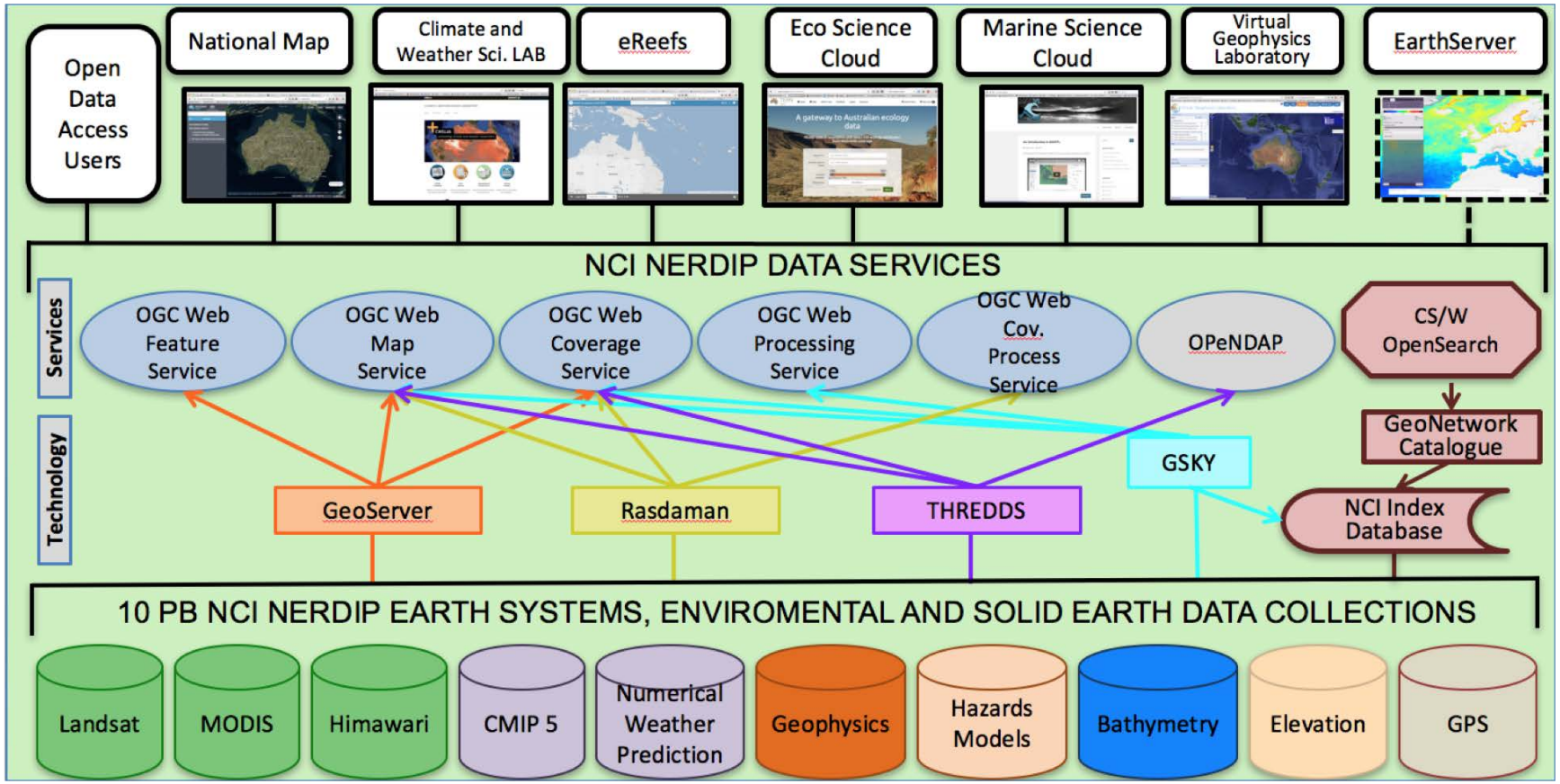








Production Data Services to access NCI datasets



- Standards-compliant
- Community-compliant
- Third-party implementations
- NCI custom implementations

NERDIP – simplified view

Compute Intensive

Virtual Laboratories

Portal views

Machine Connected

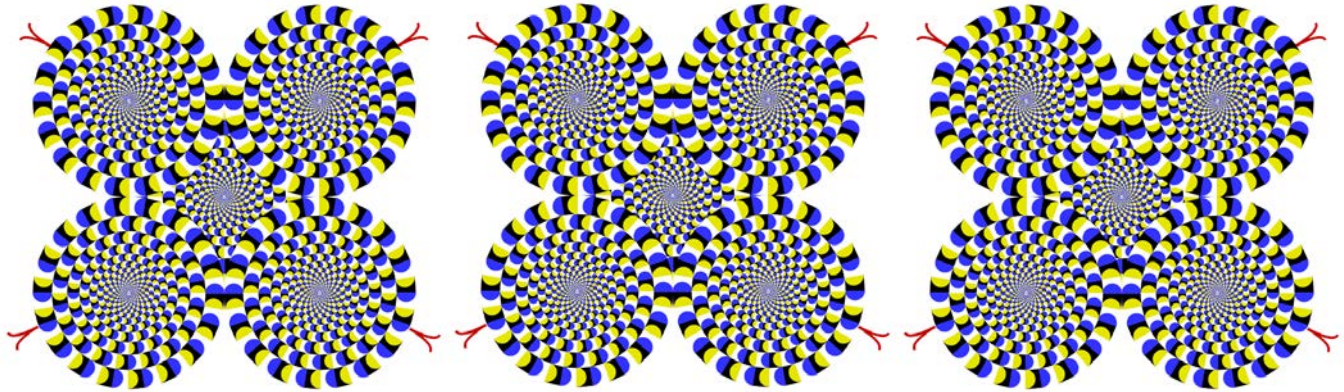
NERDIP Data Platform

Fast/Deep Data Access

Data Services

Server-side functions

Program access



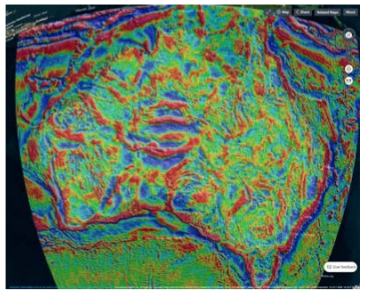
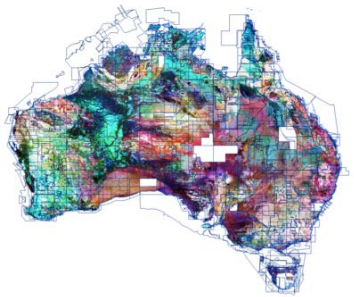
Functionality tests

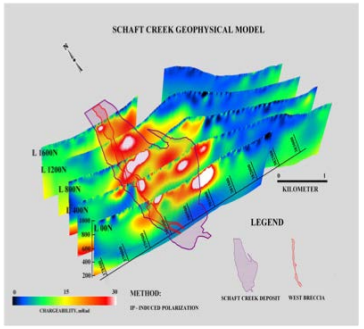
Program/Service	Test	File 1	File 2	File 3	Comments
NetCDF Utilities	ncdump (v4.3.3.1) Read netCDF file contents.	✓	✓	✓	
	NCO (v4.5.3) Read netCDF file contents.	✓	✓	✓	
GDAL Utilities (v1.11.1)	gdalinfo-1 Read netCDF file contents.	✓	✓	✓	
	gdalinfo-2 Read netCDF CRS information.	✓	✓	✓	
Data Viewers	ncview (v2.1.1) Visually inspect netCDF contents.	✓	✓	✓	Slow performance with 2-4Gb files (~mins for
	Panoply (v4.5.1) Read and plot netCDF file contents.				
THREDDS Data Server (v4.6)	File download				
	OPeNDAP (access and subsetting) Read/extract netCDF file contents.				
	Netcdf Subset Service (NCSS) Request subset of netCDF contents using spatial/temporal query.				
	Godiva WMS Viewer View netCDF file contents.				
	WMS GetMap (v1.1.1) Request netCDF file using WMS.				
	WCS GetCoverage (v1.0.0) Request netCDF file using WCS.				
Python (2.7.x) NetCDF APIs	netCDF4-python (v1.2.2) Read/extract netCDF file contents..	✓	✓	✓	
	Gdal-python (1.11.1) Read/extract netCDF file contents.	N/A	N/A	N/A	
	h5py (v2.5.0) Read/extract netCDF file contents.	N/A	N/A	N/A	
MATLAB	R2012b Read/extract netCDF file contents.	✓	✓	✓	
	R2015b Read/extract netCDF file contents.	✓	✓	✓	
	R2016a Read/extract netCDF file contents.	✓	✓	✓	
R (v3.1.0)	ncdf4 (v1.15) Read/extract netCDF file contents.	✓	✓	✓	
QGIS (v2.2.0 Valmiera)	Add data from netCDF as raster layer	N/A	N/A	N/A	
	Add data as WMS layer (served by THREDDS)	N/A	N/A	N/A	
Visualisation Tools	ParaView** (v5.0.1) Read/view netCDF file	N/A	N/A	N/A	

Overview of geophysical datasets at NCI

NCI makes available national geophysical reference datasets.
It is brought together on NCI to:

- allow high performance compute & data analysis;
- harmonise with other geophysical data types; and
- to make generally available through open data services



The data collection we are working with includes:

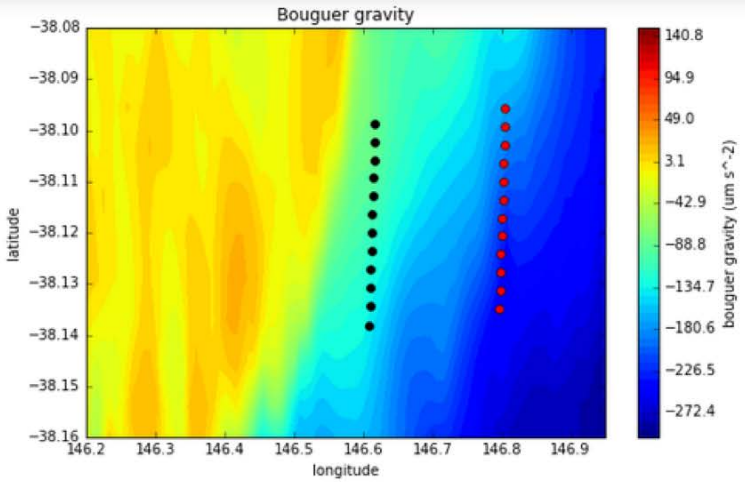
- Wide angle reflection/refraction seismic data
- Passive seismic data for undercover initiative
- Gravity
- Magnetotellurics
- Airborne – magnetic/radiometrics
- ASTER – mineral survey data

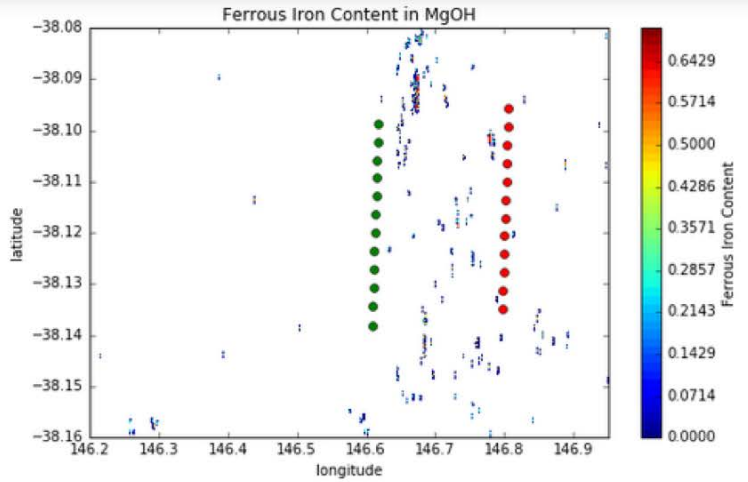
Interoperability

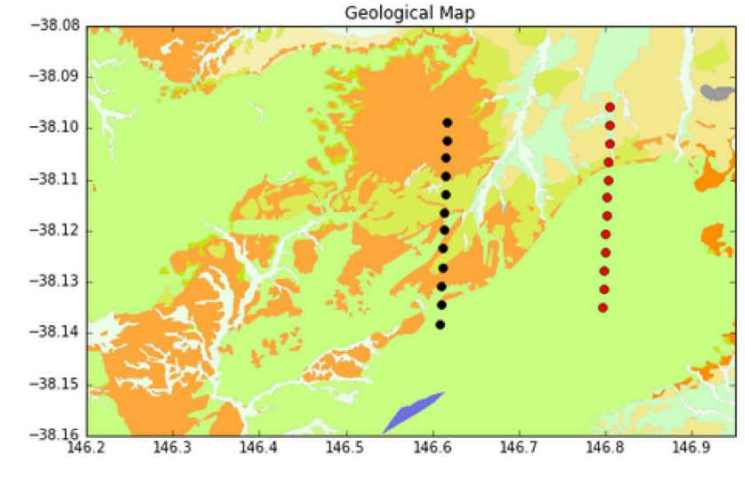
Jupyter notebook to demonstrate easy programmatic access to this data, and potential for data integration and interpolation over the Gippsland basin (Victoria, Australia) for different geophysical data types using a jupyter notebook:

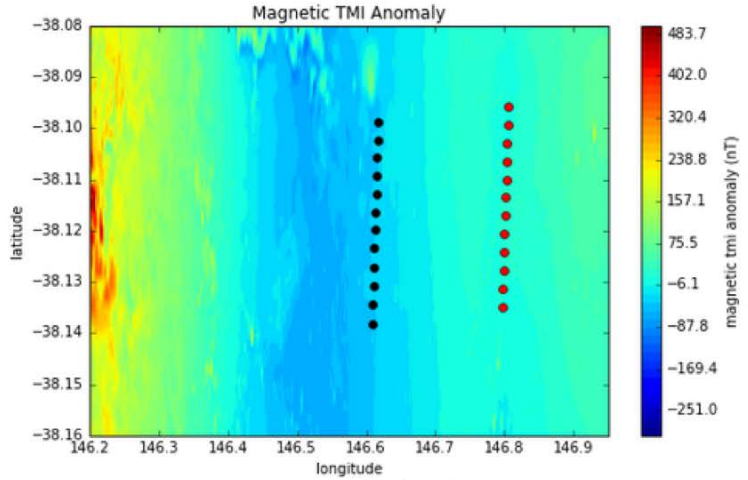
- Stacked seismic profile
- Seismic lines
- Bouguer anomaly
- Magnetic data
- ASTER mineral data
- Geological map

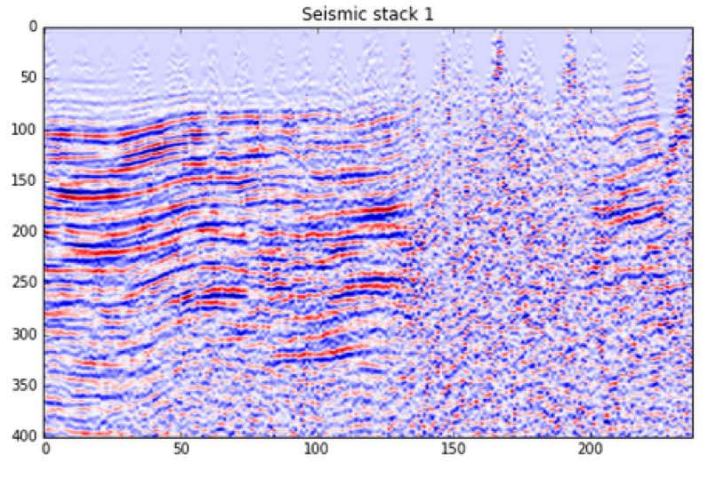
These datasets are all open to the public and have been shown to be enabled for programmatic access.

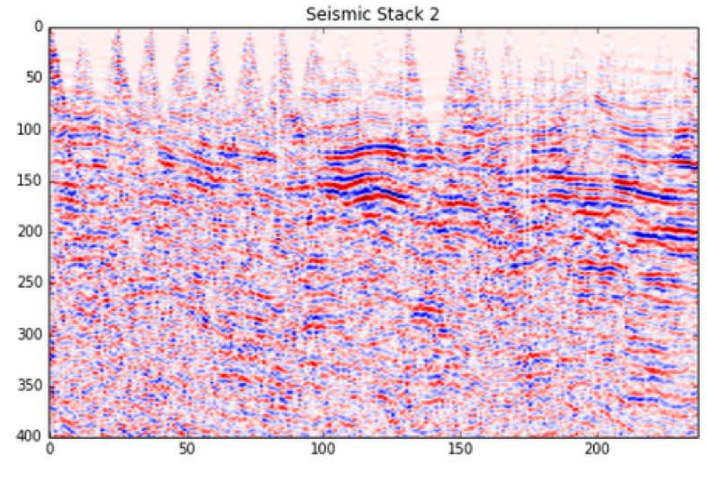












Consulting and Support

- Contact us:
  - Help trial access at NCI
  - Effectively use data
  - I have a major international project and need specific support