

**NCI**  
AUSTRALIA

# 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, known nationally and internationally for their expertise.
- Grew out of ANU Facility (mid 1980s)
- Now is a NCRIS facility – National Research Infrastructure



©EnchantedLearning.com

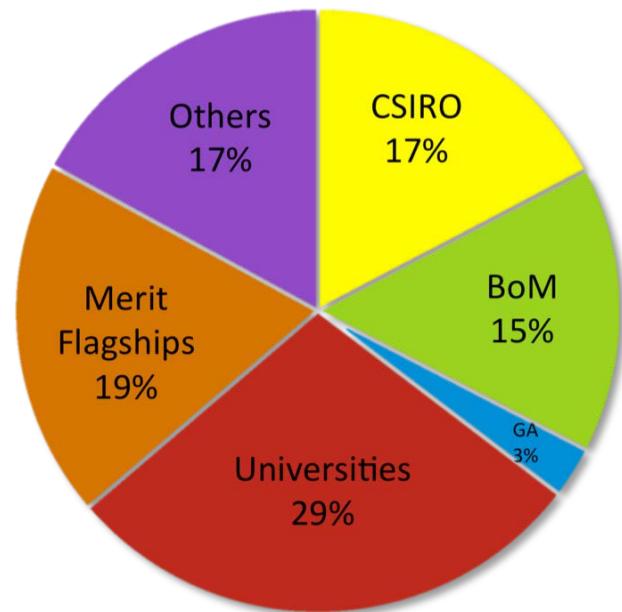
# What is NCI?

- 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)

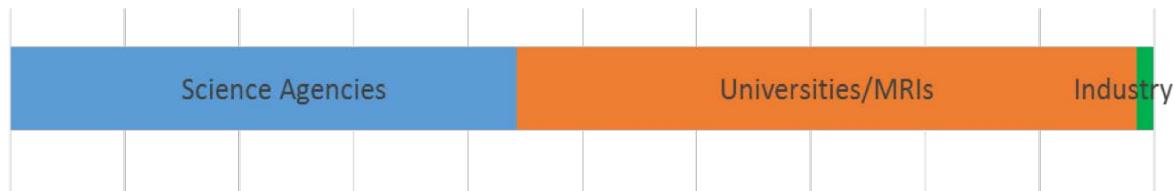


# Users of NCI

**Distribution by Share Type**

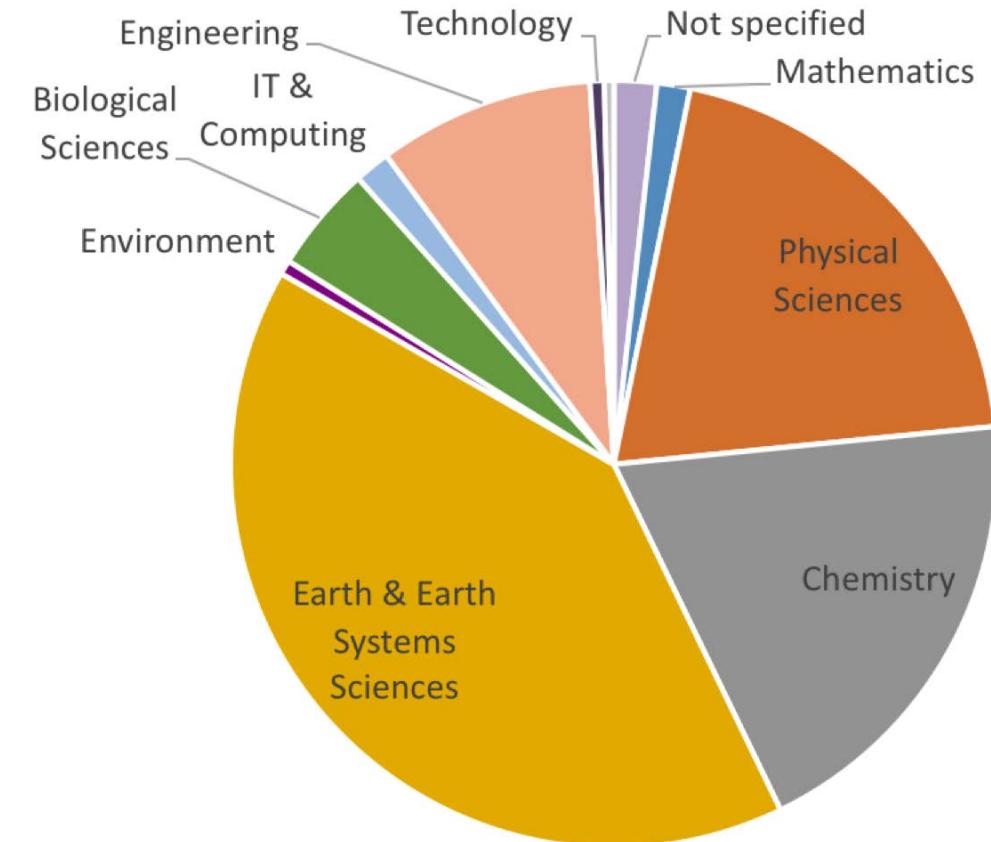


**Distribution by Organisation Type**



■ Agencies - 44.25%

■ Universities/MRIs - 54.25%

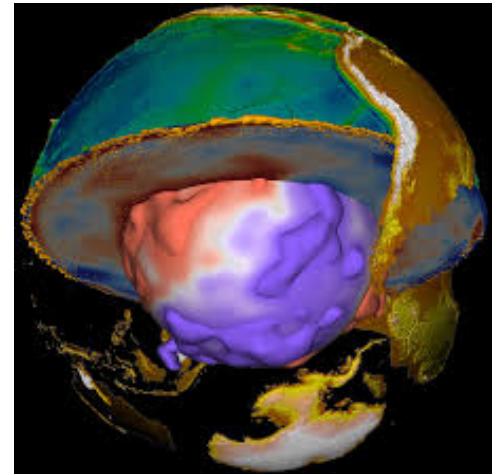


**ANU access – primarily through ANU share**  
<https://anumas.nci.org.au/2018/>

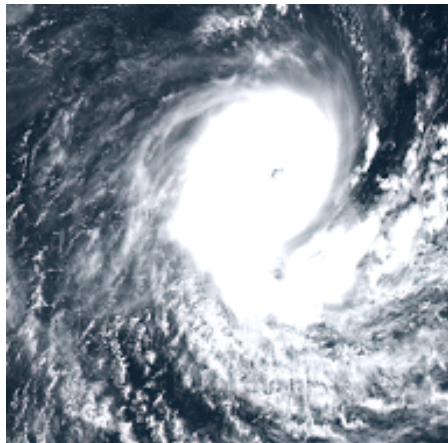
# Examples of priority areas for research at NCI

## 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

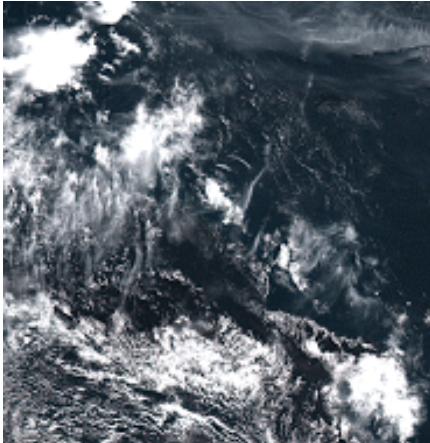


Tropical Cyclones



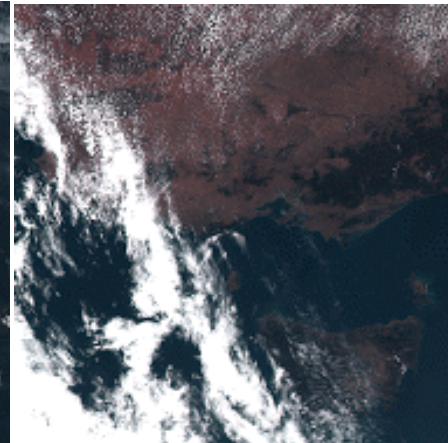
Cyclone Winston  
20-21 Feb, 2016

Volcanic Ash



Manam Eruption  
31 July, 2015

Bush Fires

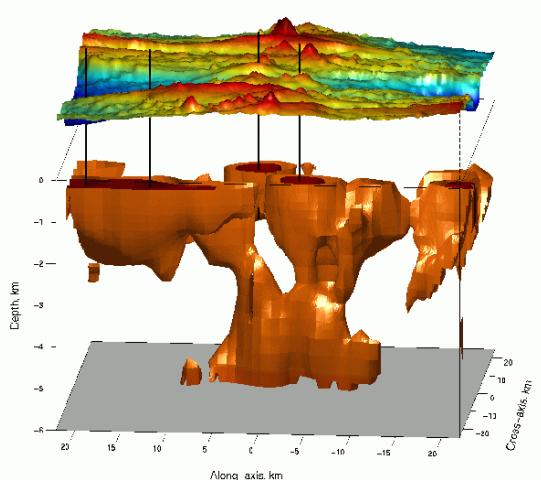


Wye Valley & Lorne Fires  
25-31 Dec, 2015

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)

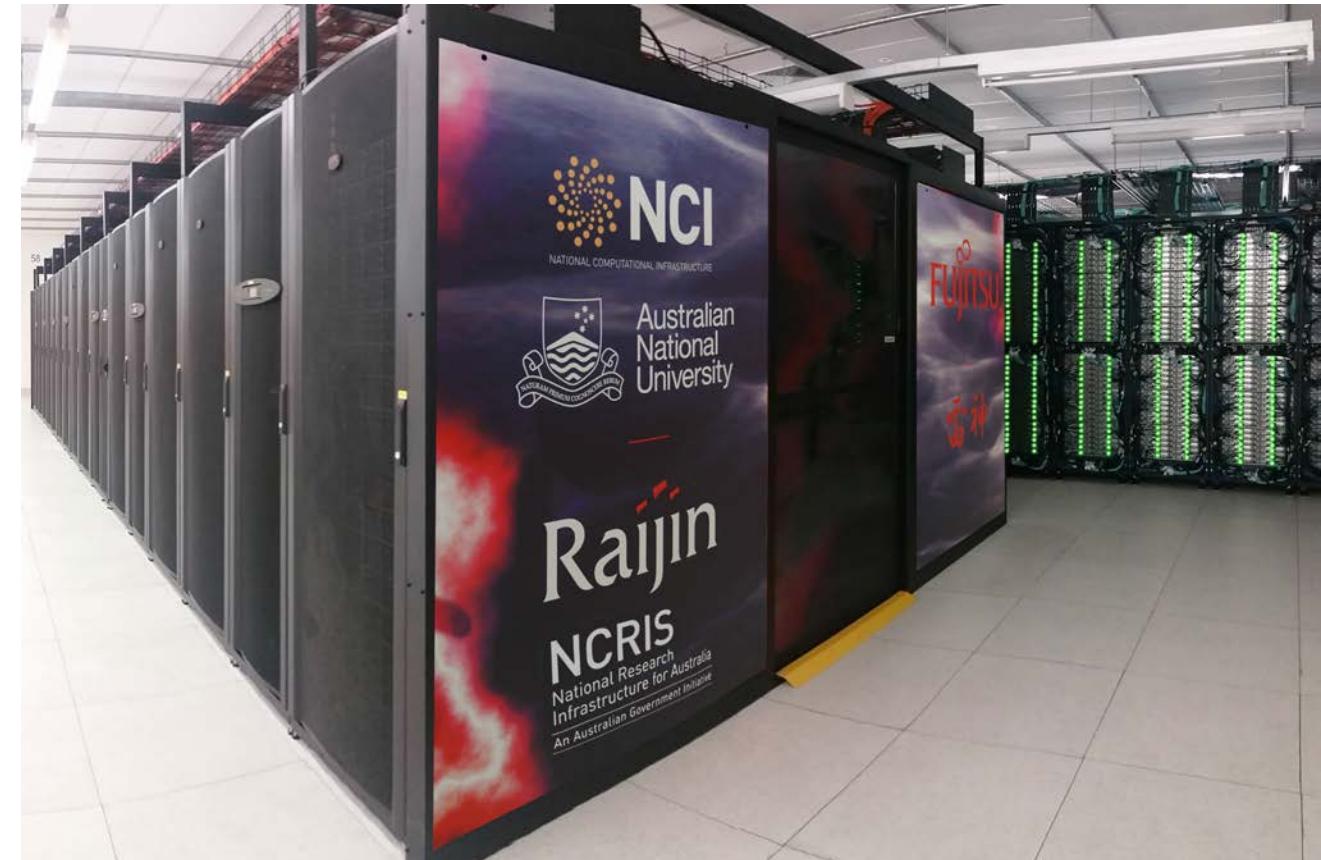
300 Terabytes of main memory

Full Fat tree interconnect (up to 100 Gb/sec)

- Hybrid FDR/EDR Mellanox Infiniband

~8 Petabytes of scratch storage

PBS batch operating system



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)

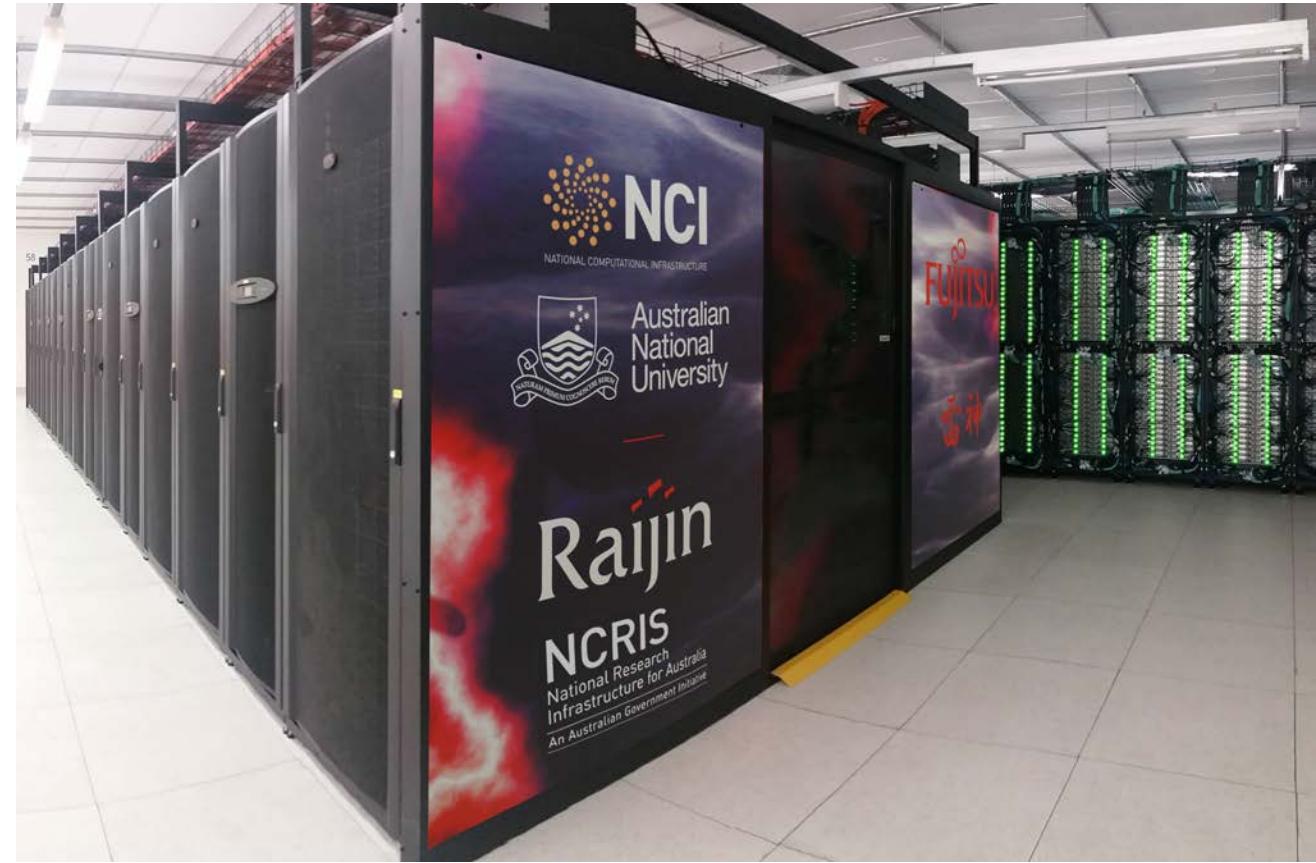
300 Terabytes of main memory

Full Fat tree interconnect (up to 100 Gb/sec)

- Hybrid FDR/EDR Mellanox Infiniband

~8 Petabytes of scratch storage

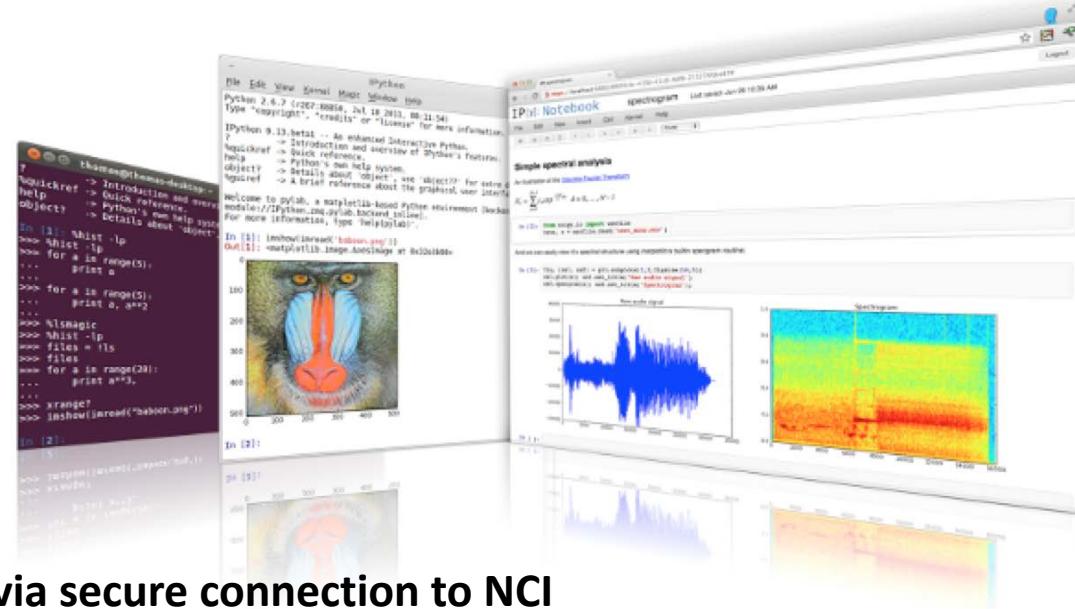
PBS batch operating system



***Significant upgrade replacement in 2019***

# VDI: interactive analysis cluster

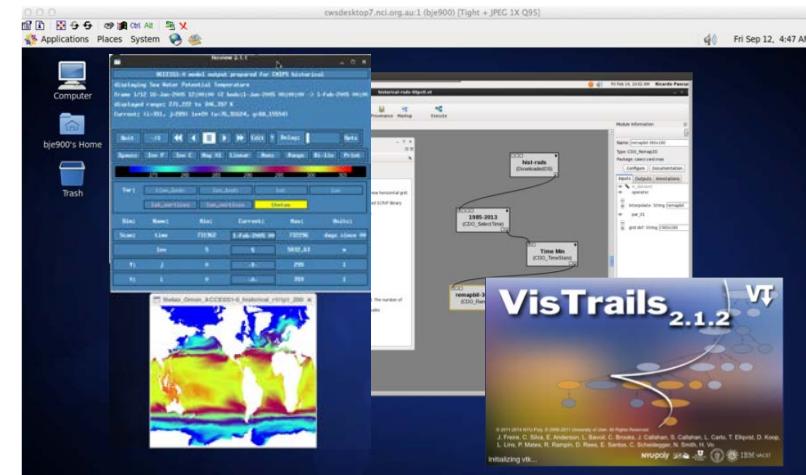
## iPython Notebooks



VDI – frictionless environment via secure connection to NCI



Strudel



**Requesting Geoscience Data**

In this notebook:

1. Manually using THREDDSDS
2. Using Python and the OPeNDAP API

The material uses Geoscience Data Services version 4.0 through NCI's THREDDSDS.

**1. Manually using THREDDSDS**

To extract data through the Web Coverage Service (WCS), the request takes the following form:

```
http://dapds00.ozgsc.nicentral.edu.au:80/service=WCS&version=1.0.0&request=GetCoverage&coverageId=ofam&bbox=-140,-60,140,0&time=2000-02-18T00:00:00Z&format=GeoTIFF
```

where red indicates required parameters that need to be defined.

GetCoverage parameters:

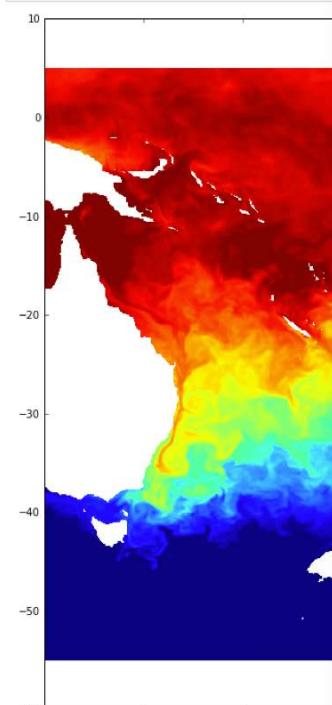
**Do the same for a smaller subset**

```
In [7]: # time slice
i = 22

T_s = ofam.variables['temp'][i, 0, 200:800]
lon_s = lon[1400:2100]
lat_s = lat[200:800]

plt.figure(figsize=(12,12))
plt.pcolormesh(lon_s, lat_s, T_s)

plt.clim(vmin=18, vmax=30)
```

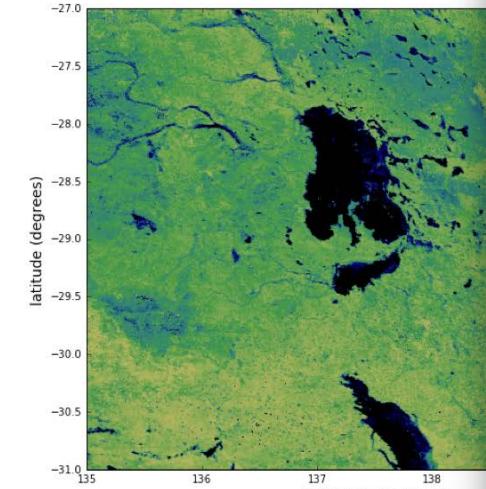


**Query a single file and view result**

```
In [6]: bbox = (135, 140, -31, -27)
lon, lat, bs, t = get_data(endpts[0], bbox)

In [7]: plt.figure(figsize=(10,10))
plt.imshow(bs, extent=bbox, cmap='gist_earth', origin='lower')
plt.xlabel('longitude (degrees)', fontsize=14)
plt.ylabel('latitude (degrees)', fontsize=14)
print "Date: ", t
```

Date: 2000-02-18



**Loop and query over the collection**

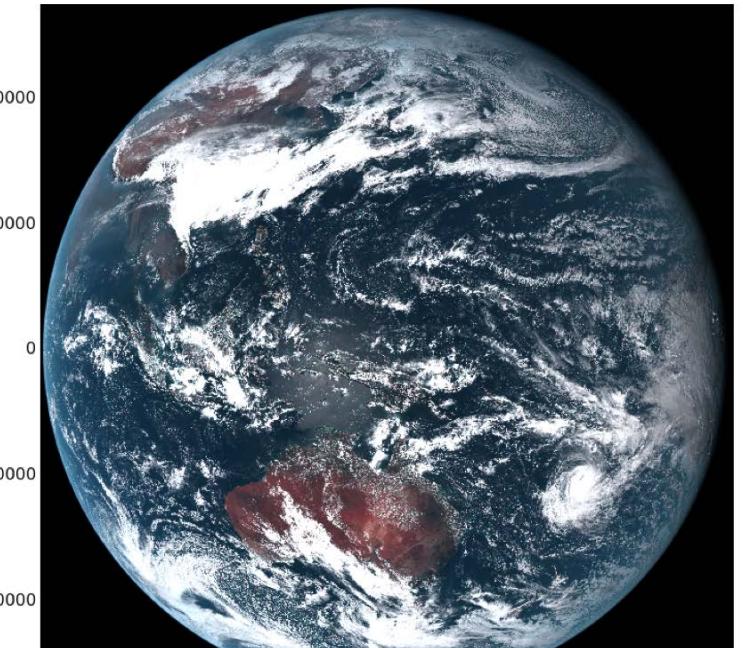
```
In [8]: bbox = (135, 140, -31, -27)
```

**Himawari 8**

```
In [13]: # Plot image
plt.figure(figsize=(12,12))
plt.imshow(rgb, extent=[x[0], x[-1], y[-1], y[0]])

# Add labels
plt.title('Himawari 8 \n', fontsize=20)

# Adjust tick mark size
plt.tick_params(labelsize=16)
```



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)

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

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>

## Metadata

### ACCESS1-3 model output prepared for CMIP5 amip

amip is an experiment of the CMIP5 - Coupled Model Intercomparison Project Phase 5 (<http://cmip-pcmdi.llnl.gov/cmip5/>). CMIP5 is meant to provide a framework for coordinated climate change experiments for the next five years and thus includes simulations for assessment in the AR5 as well as others that extend beyond the AR5.

3.3 amip (3.3 AMIP) - Version 1: AMIP (1979 - at least 2008). Impose SSTs and sea ice from observations but with other conditions as in experiment 3.2 historical.

Experiment design: [http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor\\_CMIP5\\_design.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf)  
List of output variables: [http://cmip-pcmdi.llnl.gov/cmip5/docs/standard\\_output.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/standard_output.pdf)  
Output: time series per variable in model grid spatial resolution in netCDF format  
Earth System model and the simulation information: CMIP repository

Entry name/title of data are specified according to the Data Reference Syntax ([http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5\\_data\\_reference\\_syntax.pdf](http://cmip-pcmdi.llnl.gov/cmip5/docs/cmip5_data_reference_syntax.pdf))  
as activity/product/institute/model/experiment/frequency/modelingrealm/MIP/table/ensemble  
member/version number/variable name/CMOR filename.nc

## Data and Resources

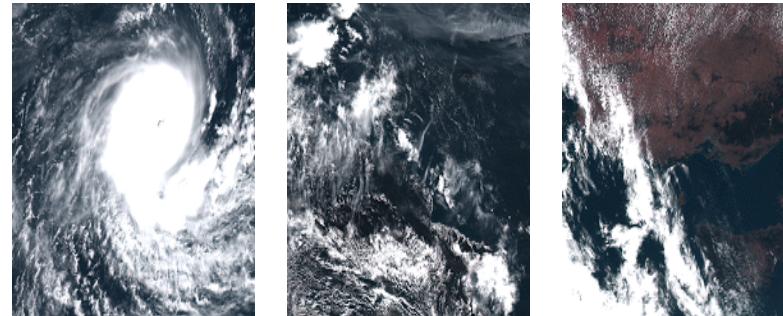
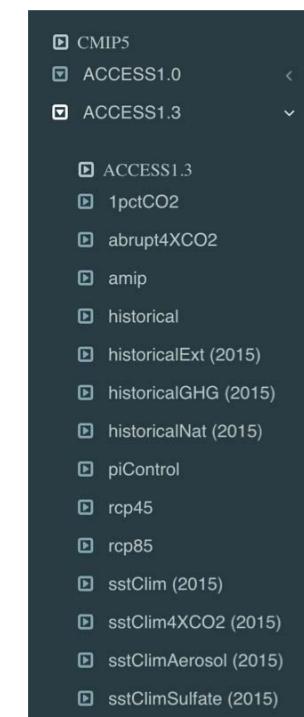
[http://pid.nci.org.au/service/TDS/15956\\_7503\\_6028\\_7510](http://pid.nci.org.au/service/TDS/15956_7503_6028_7510)

THREDDS

[http://pid.nci.org.au/dataset/15956\\_7503\\_6028\\_7510](http://pid.nci.org.au/dataset/15956_7503_6028_7510)

Geonetwork

File path on raijin : /g/data1/ua6/authoritative/IPCC/CMIP5/CSIRO-BOM/ACCESS1-3/amip



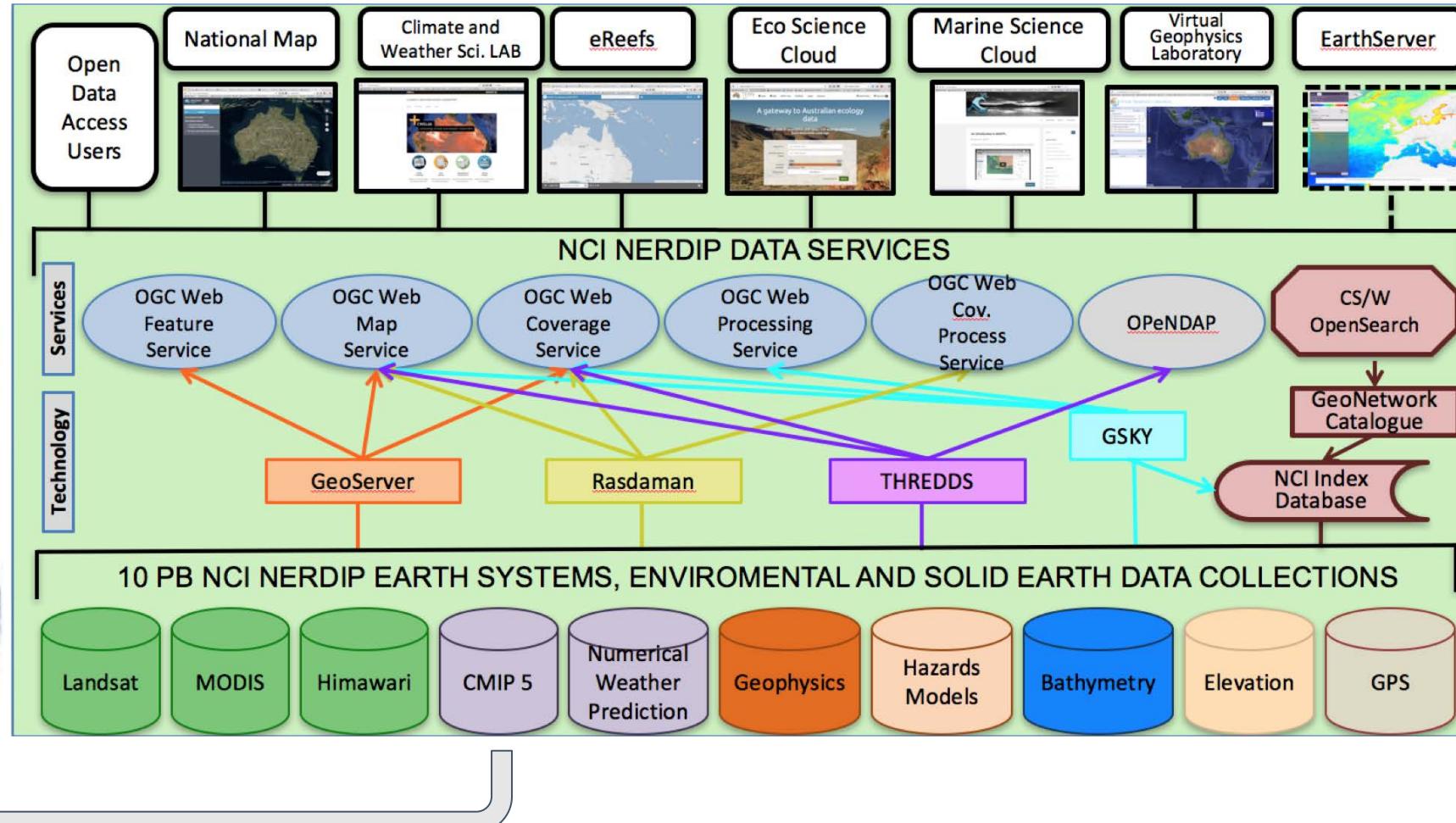
Welcome to NCI's Data Catalogue. The NCI National Research Data Collection is Australia's largest collection of research data, encompassing more than 10 PB of nationally and internationally significant datasets. A large part of our collections is publicly accessible, and touches on many different areas of science. Our collections include satellite imagery, climate and weather model outputs, and geological mapping.

Southern Sky Survey

eMAST

Global coupled climate

# Production Data Services to access NCI datasets



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

# NERDIP – simplified view

**Compute  
Intensive**



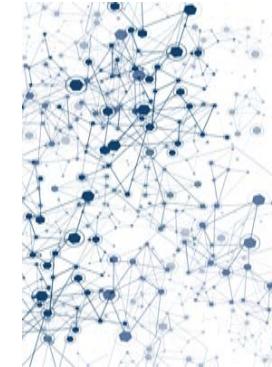
**Virtual  
Laboratories**



**Portal  
views**

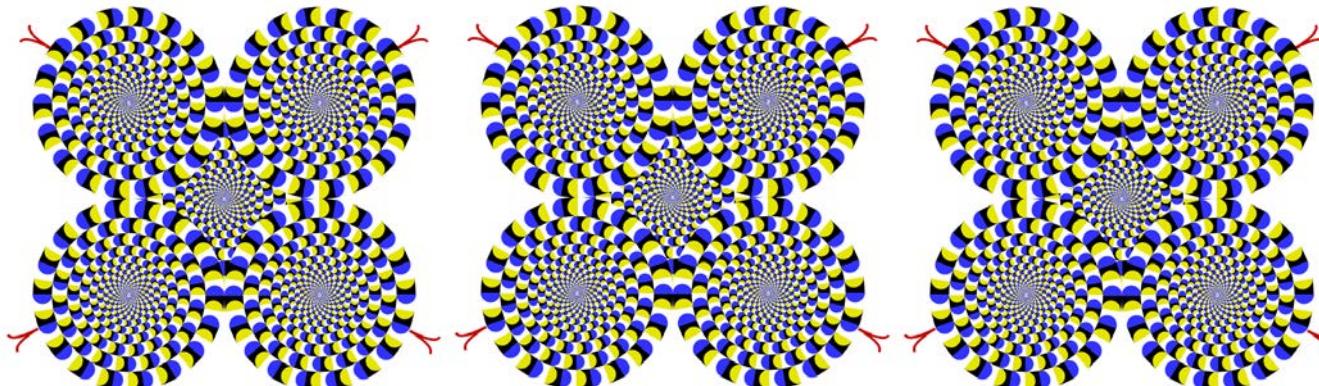


**Machine  
Connected**



## 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	<b>ncdump (v4.3.3.1)</b> Read netCDF file contents.	✓	✓	✓			
	<b>NCO (v4.5.3)</b> Read netCDF file contents.	✓	✓	✓			
GDAL Utilities (v1.11.1)	<b>gdalinfo-1</b> Read netCDF file contents.	✓	✓	✓			
	<b>gdalinfo-2</b> Read netCDF CRS information.	✓	✓	✓			
Data Viewers	<b>ncview (v2.1.1)</b> Visually inspect netCDF contents.	✓	✓	✓	Slow performance with 2-4Gb files (~mins for		
	<b>Panoply (v4.5.1)</b> Read and plot netCDF file contents.		<b>Python (2.7.x) NetCDF APIs</b>	<b>netCDF4-python (v1.1.2)</b> Read/extract netCDF file contents..	✓	✓	✓
				<b>Gdal-python (1.11.1)</b> Read/extract netCDF file contents.	N/A	N/A	N/A
				<b>h5py (v2.5.0)</b> Read/extract netCDF file contents.	N/A	N/A	N/A
	<b>THREDDS Data Server (v4.6)</b>		<b>MATLAB</b>	<b>R2012b</b> Read/extract netCDF file contents.	✓	✓	✓
				<b>R2015b</b> Read/extract netCDF file contents.	✓	✓	✓
				<b>R2016a</b> Read/extract netCDF file contents.	✓	✓	✓
	<b>OPeNDAP (access and subsetting)</b> Read/extract netCDF file contents.		<b>R (v3.1.0)</b>	<b>ncdf4 (v1.15)</b> Read/extract netCDF file contents.	✓	✓	✓
			<b>QGIS (v2.2.0 Valmiera)</b>	<b>Add data from netCDF as raster layer</b>	N/A	N/A	N/A
				<b>Add data as WMS layer (served by THREDDS)</b>	N/A	N/A	N/A
	<b>WMS GetMap (v1.1.1)</b> Request netCDF file using WMS.		<b>Visualisation Tools</b>	<b>ParaView** (v5.0.1)</b> Read/view netCDF file	N/A	N/A	N/A

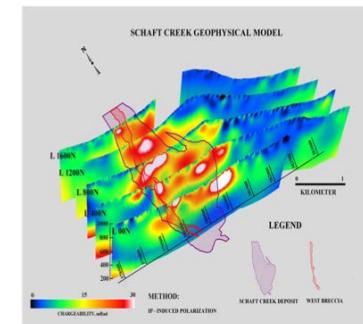
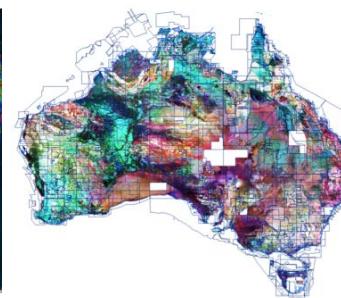
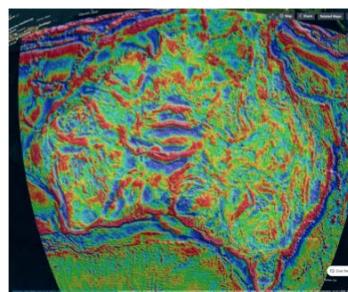


# Overview of geophysical datasets at NCI

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:

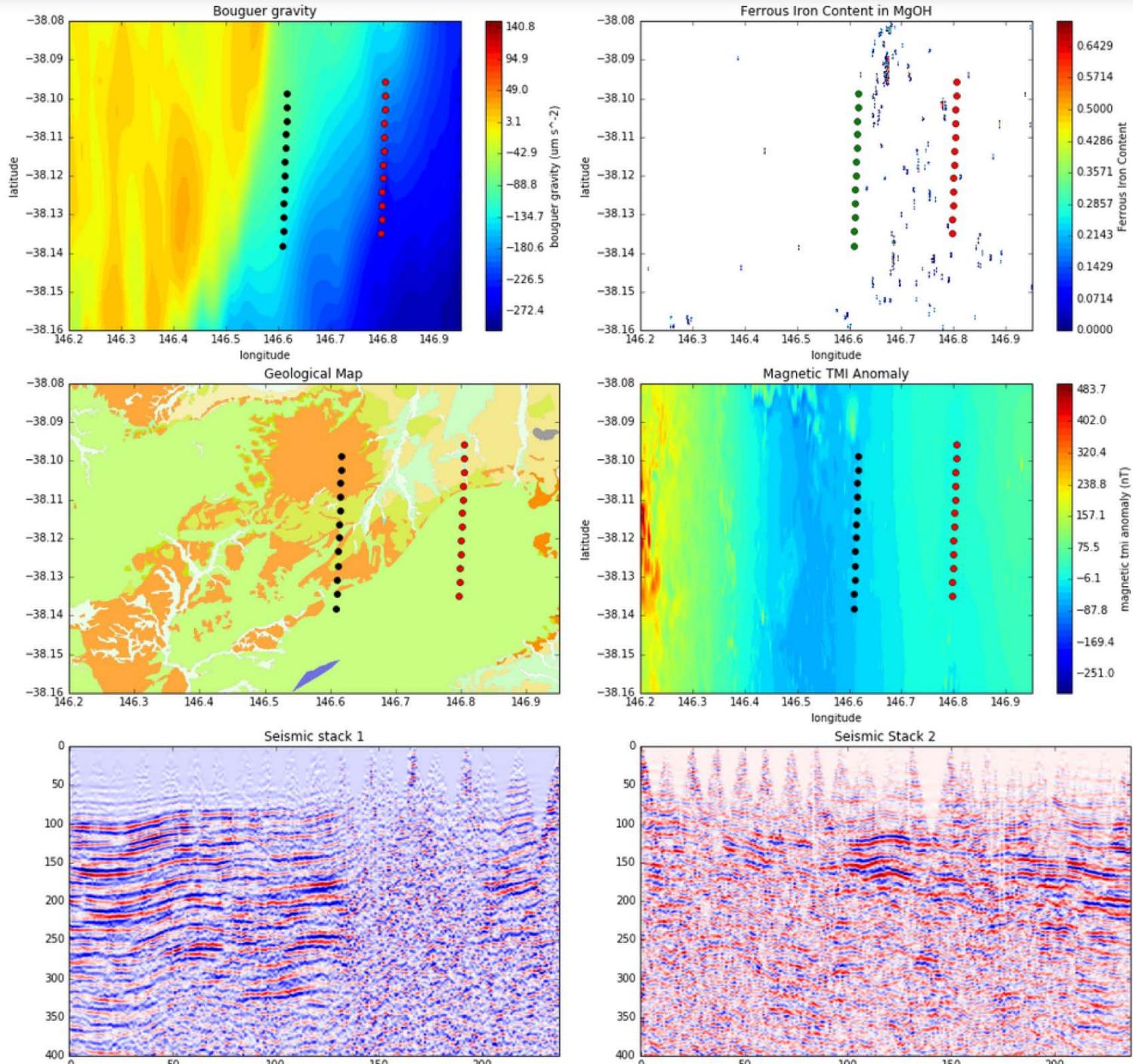
- Wide angle reflection/refraction seismic data
- Passive seismic data for undercover initiative
- Gravity
- Magnetotellurics
- Airborne – magnetic/radiometrics
- ASTER – mineral survey data

# 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

