



# Introduction to NCI's managed datasets

Jingbo Wang & Nigel Rees & Rui Yang

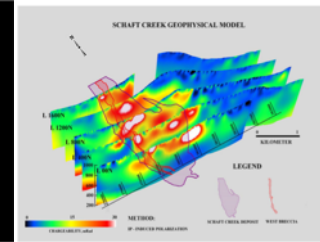
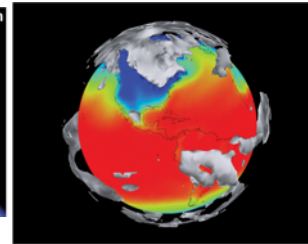
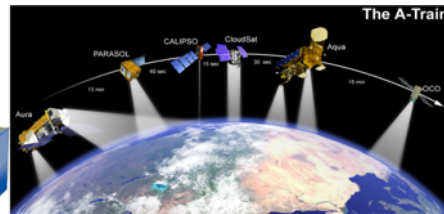
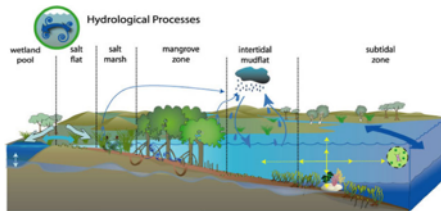
# Intro to NCI's managed FAIR datasets

Finding Data and accessing it

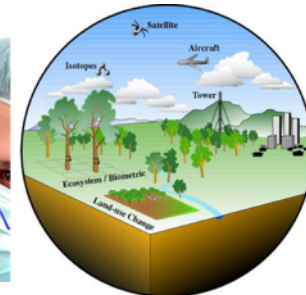
Intro to NCI's data services

Data communication channels

NCI makes available a number of large, national reference datasets. It is organised for both high performance computation & high performance data analysis, as well as making available more broadly to the research community.



- *climate and weather models*
- *satellite images*
- *bathymetry and elevation*
- *hydrology*
- *geophysics*
- *Also: optical astro, genomic and social sciences*



Data Collections	NCI project codes	Approx. Capacity
International Climate model data (e.g., CMIP5, CMIP6, CORDEX)	rr3, oi10, al33, cb20, fs38, qv56	2+ Pbytes
Bureau Weather and Climate Model data	rr4, ua4, ja4, gg6, gg8, fx1, fx3, ub3, rr8, ub7	3.5 Pbytes
Satellite data (e.g., LANDSAT, MODIS, Sentinel, Himawari, Digital Earth Australia data products)	rs0, fk4, u39, gb6, rr5	1.5 Pbytes
Ocean Models	gb6	200 Tbytes
Digital Elevation and Bathymetry	rr1	
Regional Reanalysis (BARRA)	ma05	50 Tbytes
OFES, CABLE, AGCD, CFSR	rq5, wd9, dl1, zv2	~100Tbytes
Optical Astronomy (e.g., Skymapper Southern Sky Survey)		600Tbytes

Intro to NCI's managed FAIR datasets

**Finding Data and accessing it**

Intro to NCI's data services

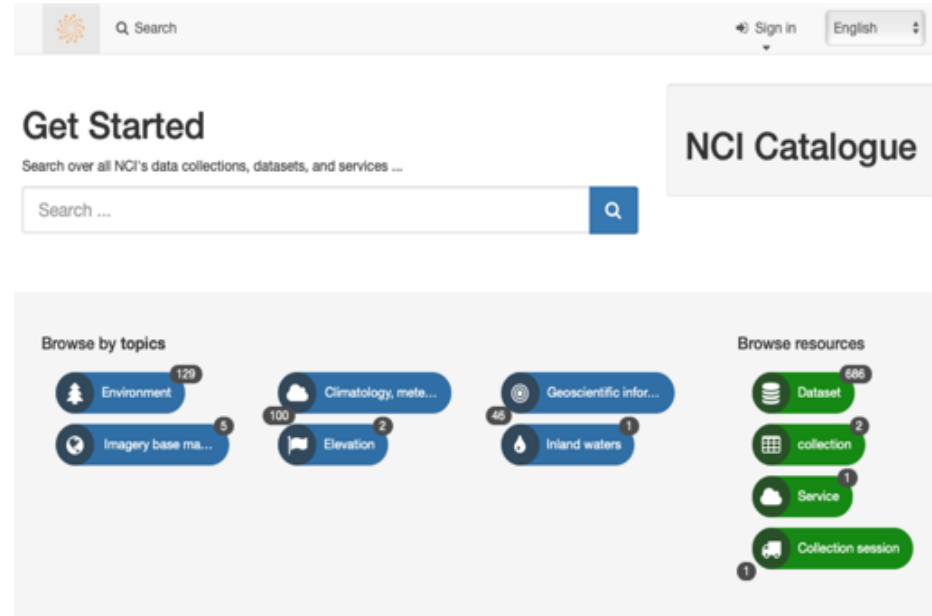
Data communication channels

- Know where to go to search for data hosted at NCI:
  - <https://geonetwork.nci.org.au>
- The options available to access data
  - Remote access options
  - In-situ on filesystem

NCI's GeoNetwork portal is our primary catalogue of data collections and is regularly updated with the latest data expanding our growing collection of datasets.

It provides an easily browsable and searchable interface to discover the datasets we manage, including descriptions of what each dataset contains, key scientific details about how it was created, and a range of access and data ownership details.

Each record has a persistent digital object identifier that can be used in publications and as a shareable link to the dataset.



The screenshot shows the NCI GeoNetwork portal interface. At the top, there is a search bar with a magnifying glass icon and the text "Q Search". To the right of the search bar are links for "Sign in" and a language dropdown menu set to "English". Below the search bar is a "Get Started" section with the text "Search over all NCI's data collections, datasets, and services ..." and a search input field with a magnifying glass icon. To the right of the search bar is a button labeled "NCI Catalogue". Below the search bar is a "Browse by topics" section with several topic buttons: "Environment" (129), "Imagery base ma..." (5), "Climatology, mete..." (100), "Elevation" (2), "Geoscientific infor..." (45), and "Inland waters" (1). To the right of the "Browse by topics" section is a "Browse resources" section with several resource buttons: "Dataset" (686), "collection" (2), "Service" (1), and "Collection session" (1).

## Get Started

Search over all NCI's data collections, datasets, and services ...

BARRA: Bureau of Meteorology high-resolution Regional Reanalysis for Australia

Typing a keywords, Geonetwork will automatically search matching records and prompt a list of records which include the keywords in the search window.

Click the record, you will find

- Abstract
- Project code and local file path
- Dataset DOI
- Data access terms and conditions
- Dataset lineages information

## BARRA: Bureau of Meteorology high-resolution Regional Reanalysis for Australia

The Bureau of Meteorology high-resolution Regional Reanalysis for Australia (BARRA) is a high-resolution multi-decadal atmospheric reanalysis. The reanalysis suite is based on the Australian Community Climate Earth-System Simulator (ACCESS) and extends 70 levels (up to 80 km) into the atmosphere. It is nested within the required boundary and/or initial conditions provided by ERA-Interim reanalysis, Operational SST and Sea Ice Analysis, and the Bureau offline soil moisture reanalysis. The region covered by the reanalysis is the Australian continent, and the surrounding region including parts of southeast Asia, New Zealand, and south to the ice edge of the Antarctic continent. About 100 parameters are available at hourly time steps at approximately 12-km resolution. For a small number of subdomains, analyses at a 1.5-km resolution are available. These have been developed using multiple convective-scale (1.5-km) downscaling analyses driven by the 12-km system.

### Record Overview

NCI project code	ma05
NCI local file path	/g/data/ma05
Catalogue record DOI	<a href="http://dx.doi.org/10.4225/41/5993927b50f53">http://dx.doi.org/10.4225/41/5993927b50f53</a>
Categories	<span style="border: 1px solid #ccc; padding: 2px;">Environment</span>
Keywords	<ul style="list-style-type: none"> <li>• <a href="#">Earth Science Reanalyses/Assimilation Models</a></li> <li>• <a href="#">Australia/New Zealand</a></li> <li>• <a href="#">1 km - &lt; 10km, 10 km - &lt; 50 km, 1 minute - &lt; 1 hour, hourly - &lt; Daily</a></li> <li>• <a href="#">AU/BOM</a></li> <li>• <a href="#">National Computational Infrastructure (NCI)</a></li> <li>• <a href="#">040107 - Meteorology</a></li> </ul>
Access constraints	<p><a href="http://dapds00.nci.org.au/thredds/fileServer/licenses/ma05_BARRA_licence_v2.pdf">http://dapds00.nci.org.au/thredds/fileServer/licenses/ma05_BARRA_licence_v2.pdf</a></p> <p>The Bureau of Meteorology licences this data under its Express Licence Agreement. Please contact the Bureau if you wish to use the data.</p>



Published collections will have a direct link to NCI's Data Services.

But what do we do from there?

(...we'll get to this bit later in the next sections!)

### Himarawi-8 and other remote and in situ observations from the Bureau of Meteorology



Collection of remotely observed and insitu observed data products from the Australian Bureau of Meteorology (Observation and Infrastructure Division, the National Meteorological and Oceanographic Centre (NMOC) and The Centre for Australian Weather and Climate Research) to support earth system modelling and ocean/marine modelling.

Within the data collection is the geostationary satellite data such as Himawari-8 for 27-01-2015 to present.

Data and metadata based on a variety of formats such as netCDF and Climate-Forecast (CF) v1.6 conventions with support for Unidata data discovery conventions, WMO BUFR format, and HDF.

More information about this collection can be found at <http://www.bom.gov.au/australia/satellite/>

#### Data Access

	<b>NCI THREDDS Data Server (all datasets)</b> <a href="http://dapds00.nci.org.au/thredds/catalog/r5/catalog.html">http://dapds00.nci.org.au/thredds/catalog/r5/catalog.html</a>	<input type="button" value="Open link"/>
	<b>NCI THREDDS Data Server (Himawari-8 data)</b> <a href="http://dapds00.nci.org.au/thredds/catalog/r5/satellite/obs/himawari8/FLDK/catalog.html">http://dapds00.nci.org.au/thredds/catalog/r5/satellite/obs/himawari8/FLDK/catalog.html</a>	<input type="button" value="Open link"/>

Q Back to search < Previous Next >

## Magnetotelluric surveys in South Australia

Updated: 20 days ago

Magnetotelluric surveys conducted in South Australia




**Completed**

**Temporal extent**  
**Period**  
 Thu Apr 05 2018 10:00:00 GMT+10C  
 10:00:00 GMT+1000

**Download and links**

[http://dapds00.nci.org.au/thredds/catalog/my80/States\\_and\\_Territories/SA/catalog.html](http://dapds00.nci.org.au/thredds/catalog/my80/States_and_Territories/SA/catalog.html)

**Associated resources**

 <p><b>Broadband and long-period MT survey acquired along the southern part of the GOMA (Gawler Craton, Officer Basin, Musgrave Province, Amadeus Basin) deep seismic reflection transect (GA08-OM1) from Tarcoola to near Coober Pedy in South Australia (beside the Adelaide-Darwin railway line), collected in December 2008</b></p> <p>In December 2008, broadband and long-period magnetotelluric (MT) data were acquired by Geoscience Australia (GA) along the southern part of the GOMA...</p>	<input type="button" value="Child record"/>
 <p><b>Broadband MT survey in the Renmark Trough, collected in 2009.</b></p> <p>There have been numerous geophysical investigations into the Renmark Trough. Gravity data was collected in the 1960s and mid-1980s. A large seismic...</p>	<input type="button" value="Child record"/>
 <p><b>Crustal scale magnetotelluric data survey along the 08 GA-C1 seismic line in the Curnamona Province in South Australia, collected in 2008.</b></p>	<input type="button" value="Child record"/>

**Provided by**

**Share on social sites**

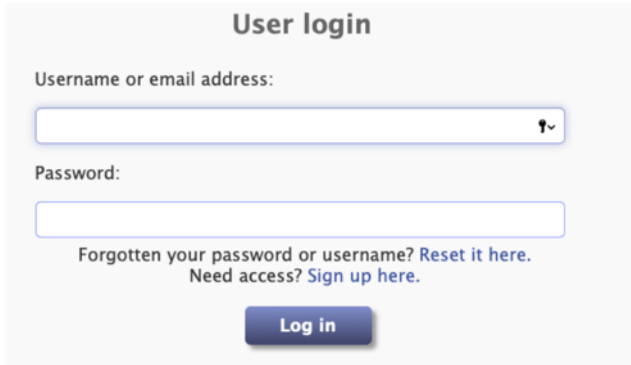
Rating: ☆☆☆

## Four common ways to access data at NCI:

1. Raijin/Gadi – membership required
2. VDI – membership required
3. Web services – free open\*
  - Data services (e.g. THREDDS, GSKY)
4. Specific data portals
  - (e.g., ESGF, SARA, Auscope, Skymapper, National Map)

\*with a few exceptions depending on the datasets.

If the data has access restrictions (the terms and conditions) then users need to be granted read-only access to the data. This can be done through Mancini <http://my.nci.org.au>



**User login**

Username or email address:

Password:

Forgotten your password or username? [Reset it here.](#)  
Need access? [Sign up here.](#)

**Log in**

## New user:

sign up (5-step registration)

## NCI user:

Projects and groups → Find project or group → type project code or keywords → select project → click “Join” → read and agree terms and conditions → submit membership request → wait for project CI’s decision

## Data Access

Register for local access

<https://my.nci.org.au/mancini/project/ua8/join>

NCI local file path

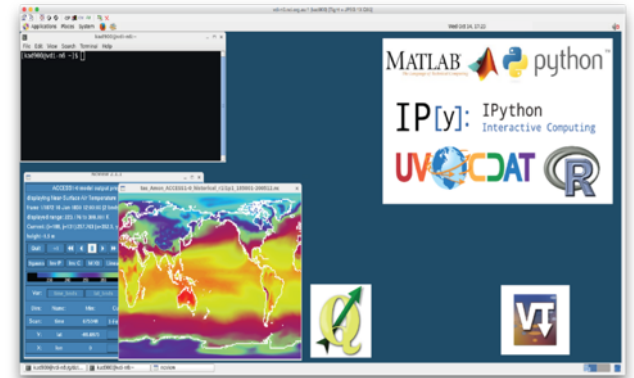
/g/data1/ua8/ARCCSS\_Data/ACCESSFire/v1-0

## What is VDI?

NCI's Virtual Desktop Infrastructure (VDI) is a complementary service to the NCI supercomputer, providing an interactive scientific desktop environment loaded with an extensive library of software packages and giving access to the datasets within NCI's internal high-speed network.

- Jupyter Notebook
- Panoply
- Paraview
- Matlab
- DASK
- QGIS
- ...

For more information, please read the [VDI User Guide](#).



Step 1: you need to install VDI on your local machine. See instruction [here](#).

Step 2: login VDI using your NCI account. Sign up [here](#) if you don't have an account.

Step 3: Once you login, open a terminal, load a few python modules.

Step 4: You can find instructions on how to set up python environment and use jupyter notebooks on VDI [here](#) (see session: How to run Jupyter notebooks).

Step 5: stay tune for the upcoming session about VDI in the near future.

Intro to NCI's managed FAIR datasets

Finding Data and accessing it

**Intro to NCI's data services**

Data communication channels

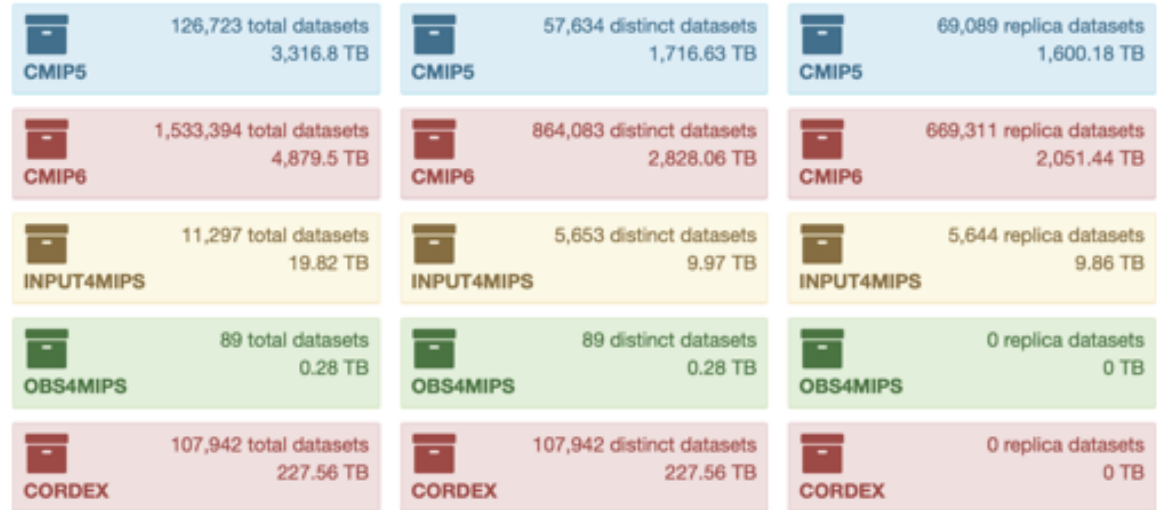
- Scale of data size and structure:  
too many/big datasets  
too complicated to drill down
- data transfers limit  
network/cost
- compute resources not available  
somewhere else  
MPI/GPU
- Performance  
I/O, Memory issue
- External push on open data for  
sharing  
Funding agency  
Publisher  
Transparency  
Need to be validated

## Published data over the federation

### All projects



### Top projects



<http://esgf-ui.cmcc.it:8080/esgf-dashboard-ui/pages/index.html>



For many of our data collections, you have the option to access data remotely through our data services.

This allows you to access data on the file without needing to copy/rsync data across locally.

Particularly useful when accessing smaller subsets of larger collections.

Data-As-A-Service to provide user:

- F.A.I.R. datasets
- Insitu access
- Seamless integration
- Co-locate HPC

Data  
download era

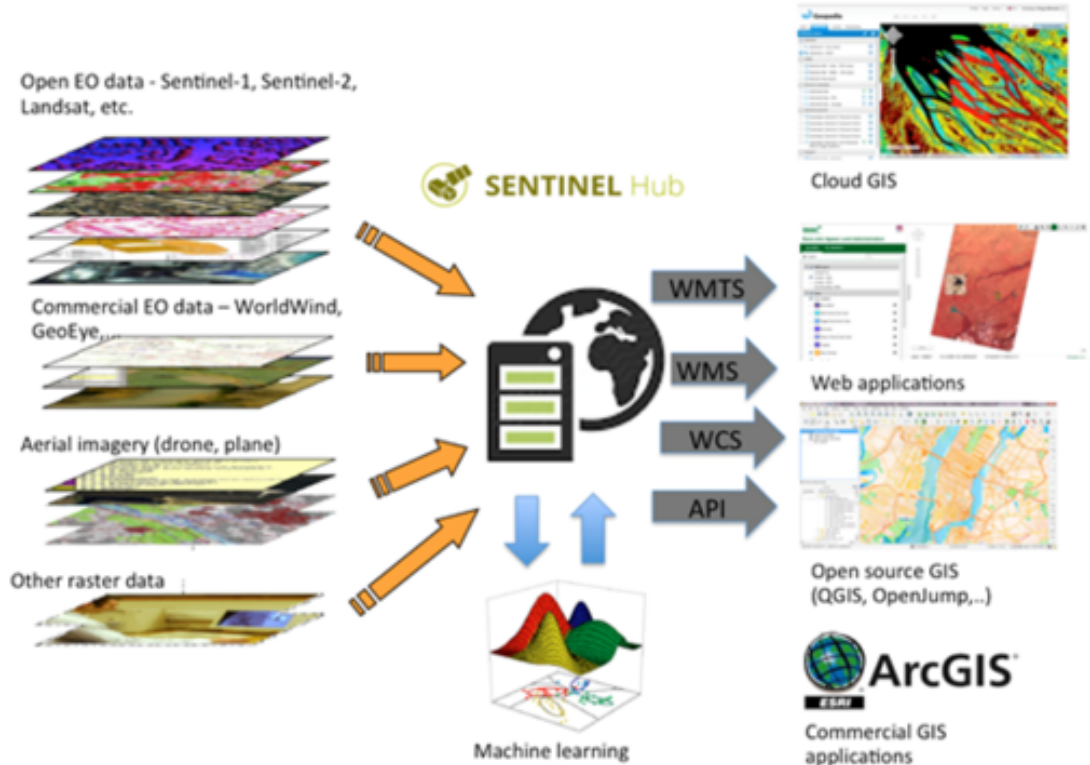
Cloud access  
and ready to  
use era

## THREDDS:

- OGC WMS
- OGC WCS
- OPeNDAP
- Subsetting
- Aggregation
- visualization

## GSKY:

- WMS
- WCS
- WPS
- WCPS



Look at what you have and find what I need → Ask the system to give me what I want

- data transfers limit

- Not needed anymore

Not possible → Totally possible

sible

I/O, Memory issue

- External push on open data for sharing

- Possible now,  
Quick

Disconnected → F.A.I.R

Need to be validated

## What is THREDDS?

THREDDS (Thematic Realtime Environmental Distributed Data Services) data server (TDS) developed by Unidata (UCAR) allows for browsing and accessing of data (as well as metadata)

Name	Description
OPeNDAP (DAP2)	Protocol enabling data access and subsetting through the web
NetCDF Subset Service (NCSS)	Web service for subsetting files that can be read by the netCDF java library
Web Map Service (WMS)	OGC web service for requesting static images of data
Web Coverage Service (WCS)	OGC web service for requesting data in some output format
Godiva Data Viewer	Tool for simple visualisation of data
HTTP File Download	Direct downloading

**Brief overview of these services:**

Name	Description
OPeNDAP (DAP2)	Protocol enabling data access and subsetting through the
NetCDF Subset Service (NCSS)	Web service for subsetting files that can be read by the netCDF java library
Web Map Service (WMS)	OGC web service for requesting static images of data
Web Coverage Service (WCS)	OGC web service for requesting data in some output format
Godiva Data Viewer	Tool for simple visualisation of data
HTTP File Download	Direct downloading

**Data subsetting options**

# OPeNDAP

## Allows for:

Remote access and subsetting. In many tools, the OPeNDAP data URL can also be used in the same manner as in-situ files.

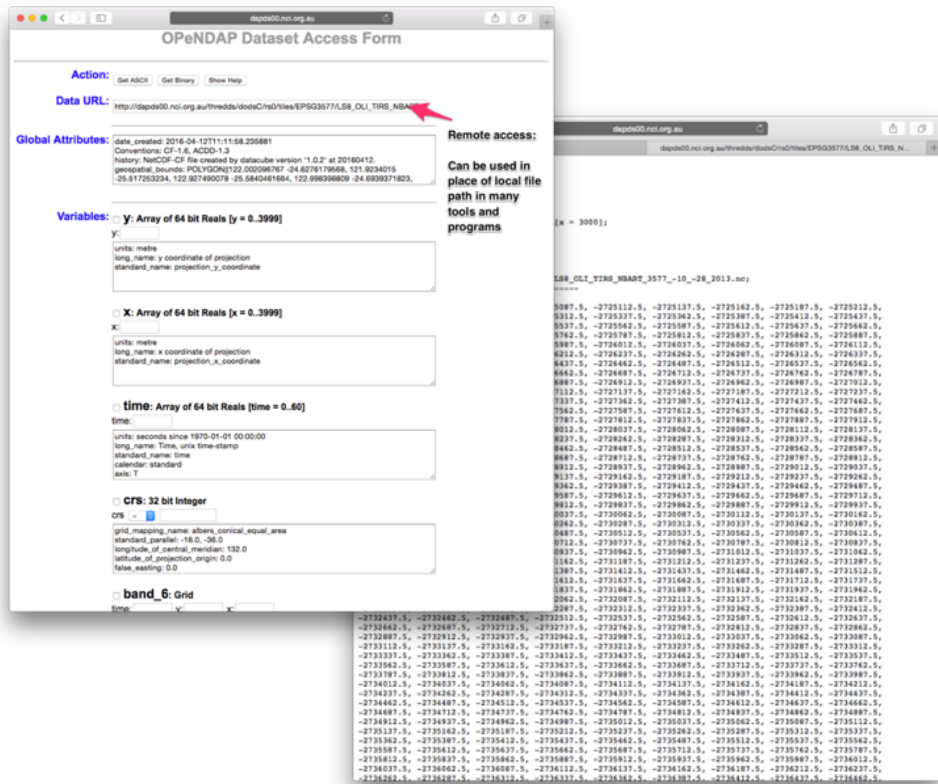
## Ideal for:

Remote use with tools and software when working with small subsets of large datasets.

When there is no pressing need to download data.

## Possible cons:

Everything is index based.



The screenshot shows the OPeNDAP Dataset Access Form in a web browser. The 'Data URL' field contains a long URL: `http://dep00.nci.org.au/thredds/ioda/Crs0/lines/EP5G3677L58_DLI_TIRS_NBAN..._2013.nc`. A red arrow points to this URL. Below the URL, there are sections for 'Global Attributes' and 'Variables'. The 'Variables' section lists several variables, including 'y: Array of 64 bit Reals [y = 0.3999]', 'x: Array of 64 bit Reals [x = 0.3999]', 'time: Array of 64 bit Reals [time = 0.60]', 'CRS: 32 bit Integer', and 'band\_6: Grid'. A 'Remote access:' box on the right states: 'Can be used in place of local file path in many tools and programs'. The bottom of the screenshot shows a preview of the data as a grid of numerical values.

## Netcdf Subset Service (NCSS)

### Allows for:

Data subsetting of large files based on spatial and/or temporal queries.

### Ideal when:

Needing to extract data for specific locations or time. Unlike OPENDAP, which is all index based, NCSS will do the work to convert the spatial/temporal query into index locations.

### Possible cons:

Downloading data locally (when using through the THREDDIS interface).

The image displays two screenshots of the NetCDF Subset Service (NCSS) web interface. The top screenshot shows the 'Grid as Point Dataset' view, where a list of variables is displayed, and a spatial subset map is visible. The bottom screenshot shows the 'Gridded Dataset' view, where a list of variables is displayed, and a 'Choose Lat/Lon Location' section is visible, allowing users to specify latitude and longitude coordinates. Green arrows point from text labels to these specific UI elements: 'variables' points to the variable list, 'latitude, longitude' points to the location input fields, 'time extents' points to the time range selection, and 'output format' points to the output format dropdown.

**Brief overview of these services:**

Name	Description
OPeNDAP (DAP2)	Protocol enabling data access and subsetting through the web
NetCDF Subset Service (NCSS)	Web service for subsetting files that can be read by the netCDF java library
Web Map Service (WMS)	OGC <b>OGC Web Services</b> data
Web Coverage Service (WCS)	OGC web service for requesting data in some output format
Godiva Data Viewer	Tool for simple visualisation of data
HTTP File Download	Direct downloading



## Web Map Service (WMS)

**Allows for:**

Allows for generation of static images for data previewing

**Ideal when:**

Ideal when further analysis is not required, or when data are for display on web-based mapping systems (e.g., web portals)

## Web Coverage Service (WCS)

**Allows for:**

Allows for requesting raw data from remote sources as TIFF or NetCDF files

**Ideal when:**

Ideal when collecting useable subsets of remote datasets for further analysis

**Brief overview of these services:**

Name	Description
OPeNDAP (DAP2)	Protocol enabling data access and subsetting through the web
Netcdf Subset Service (NCSS)	Web service for subsetting files that can be read by the netCDF java library
Web Map Service (WMS)	OGC web service for requesting static images of data
Web Coverage Service (WCS)	OGC web service for requesting data in some output format
Godiva Data Viewer	<div data-bbox="780 784 1553 912" style="background-color: #1a3d54; color: white; text-align: center; padding: 10px; font-size: 1.2em;">Visualising data</div>
HTTP File Download	Direct downloading

## Godiva: ncWMS data viewer

- Ability to browse through time and space
- Make animations
- Export KMZ format

### Allows for:

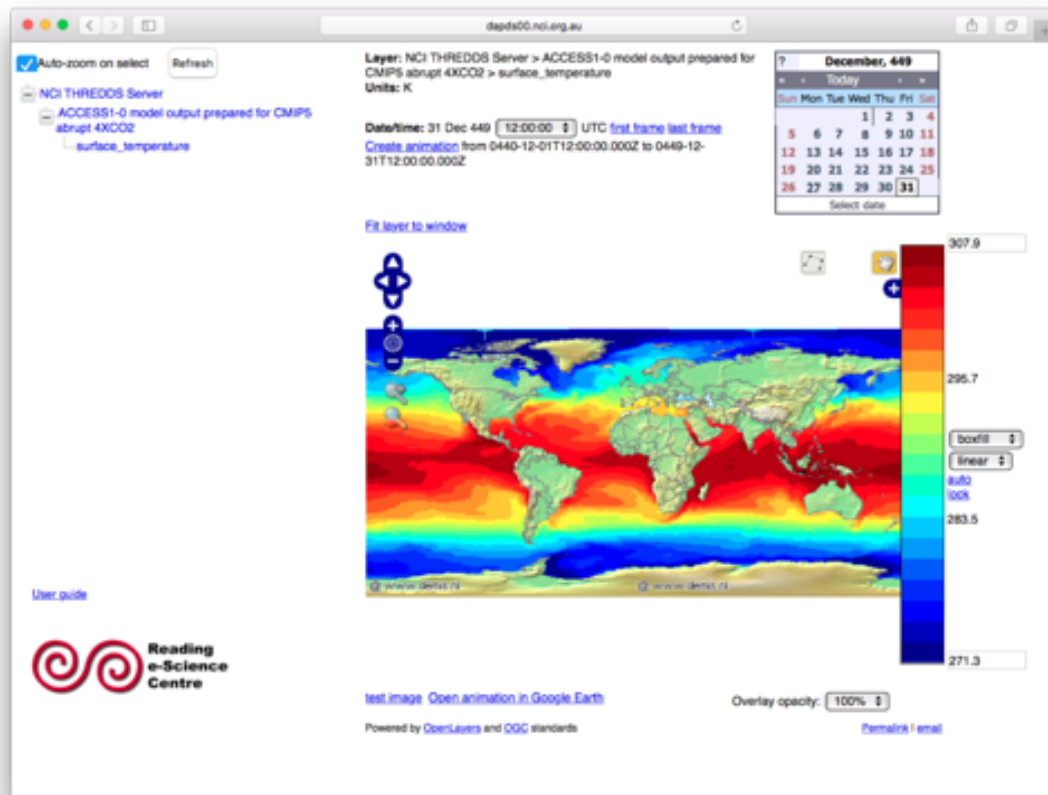
Quick visualisation from your web browser.

### Ideal when:

Inspecting data before accessing through other services.

### Con:

Not all coordinate projections are supported.



## Brief overview of these services:

Name	Description
OPeNDAP (DAP2)	Protocol enabling data access and subsetting through the web
Netcdf Subset Service (NCSS)	Web service for subsetting files that can be read by the netCDF java library
Web Map Service (WMS)	OGC web service for requesting static images of data
Web Coverage Service (WCS)	OGC web service for requesting data in some output format
Cedive Data Viewer	Tool for simple visualisation of data
HTTP File Download	Direct download of data files

**Downloading full file**

## Data Download

### Allows:

Users to directly download files to individual workstations.

### Ideal when:

Direct access or other data services are not sufficient in particular workflows.

### Possible cons:

Downloading data, particularly if you only need to work with small subset of larger collection. Volumes may be very large. Creating copies can introduce versioning problems.

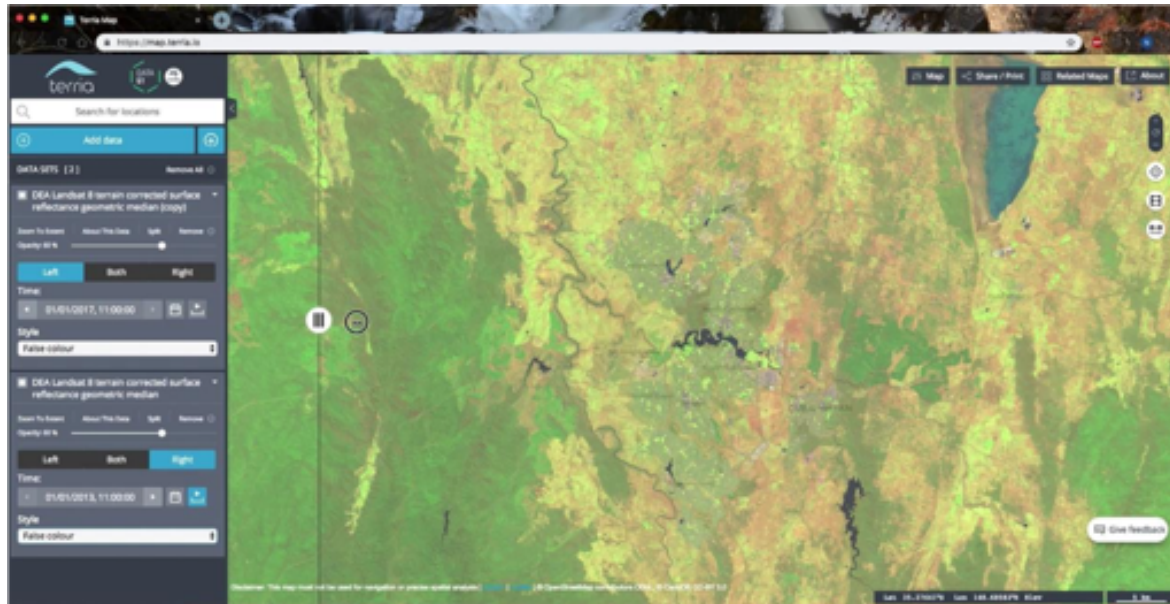


Image:

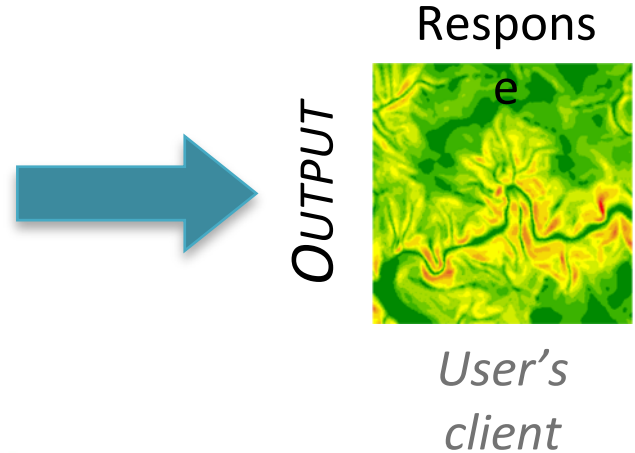
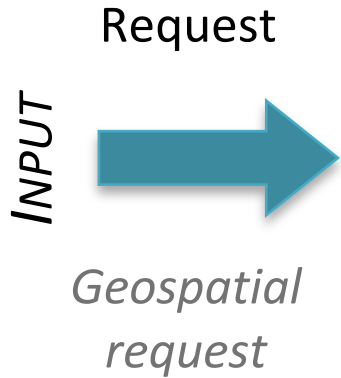
[https://www.google.com.au/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKFwjOrJi50nQAhVJlopOKHab7CwsQjRwIBw&url=http%3A%2F%2F4bg.com.au%2Fdownload-f4hg-documents-tools%2F&psig=AFQICNHIE8Dk8RwN\\_uG2NHnf6dsGS3e2tA&ust=1478161379040807](https://www.google.com.au/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=0ahUKFwjOrJi50nQAhVJlopOKHab7CwsQjRwIBw&url=http%3A%2F%2F4bg.com.au%2Fdownload-f4hg-documents-tools%2F&psig=AFQICNHIE8Dk8RwN_uG2NHnf6dsGS3e2tA&ust=1478161379040807)

## What is a data portal?

*A list of datasets with pointers to how those datasets can be accessed.  
(<https://blog.ldodds.com/2015/10/13/what-is-a-data-portal/>)*

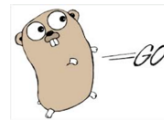


## GSKY



### FEATURES

- Distributed
- Scalable
- Concurrent



PostGIS



kubernetes



GDAL - Geospatial Data Abstraction Library



The Apollo moon landings were supported by a computer interface known as DSKY. Astronauts could input data and commands into the keypad and see the results returned on an electronic display. In much the same way, GSKY is an interface that allows human manipulation of deeply buried geospatial data. Using GSKY, a user can make complex requests and see the results in their web browser in near real-time.

**GSKY** cannot navigate its users to the moon – it can, however, help us understand it.



## **GSKY responds to Open Geospatial Consortium (OGC) API over http protocol:**

- Web Map Service (for displaying the images on the map server)
- Web Coverage Service (for delivering the actual data as “coverages” - independent of the underlying storage format or files)
- Web Processing service

## **GSKY allows:**

- Performant aggregations, subsetting, subsampling, polygon/pencil/pixel drills
- Execution of on-the-fly data transformations, re-projections and other algorithms
- Deal with very large files

## **GSKY is implemented using**

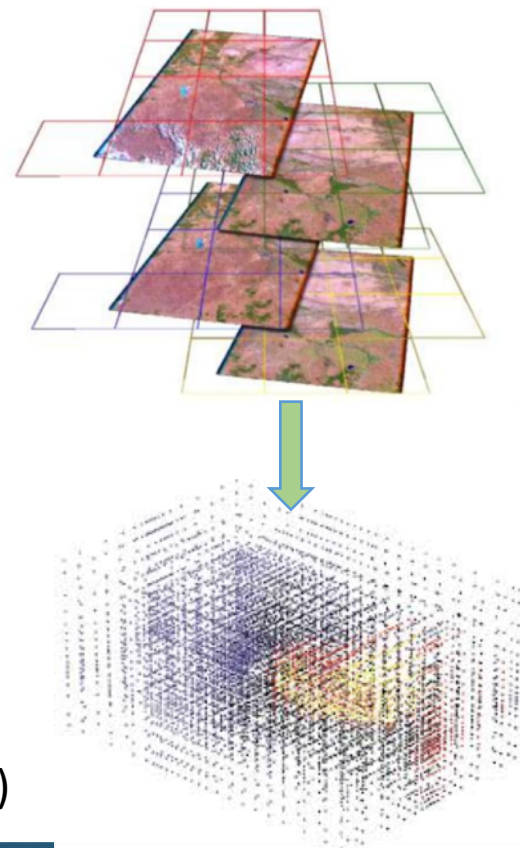
- Rich metadata server for data query e.g., spatial, temporal, other physical variables
- Clustered backend workers – high performance I/O and scale-out server-side compute

**MAS:** Integrated with GSKY and provides a backend abstraction over the data

- It identifies individual data objects (datasets, variables, spatial and temporal extents).
- Can be shared by different geospatial collections or by splitting collections into non-overlapping geographical extents.

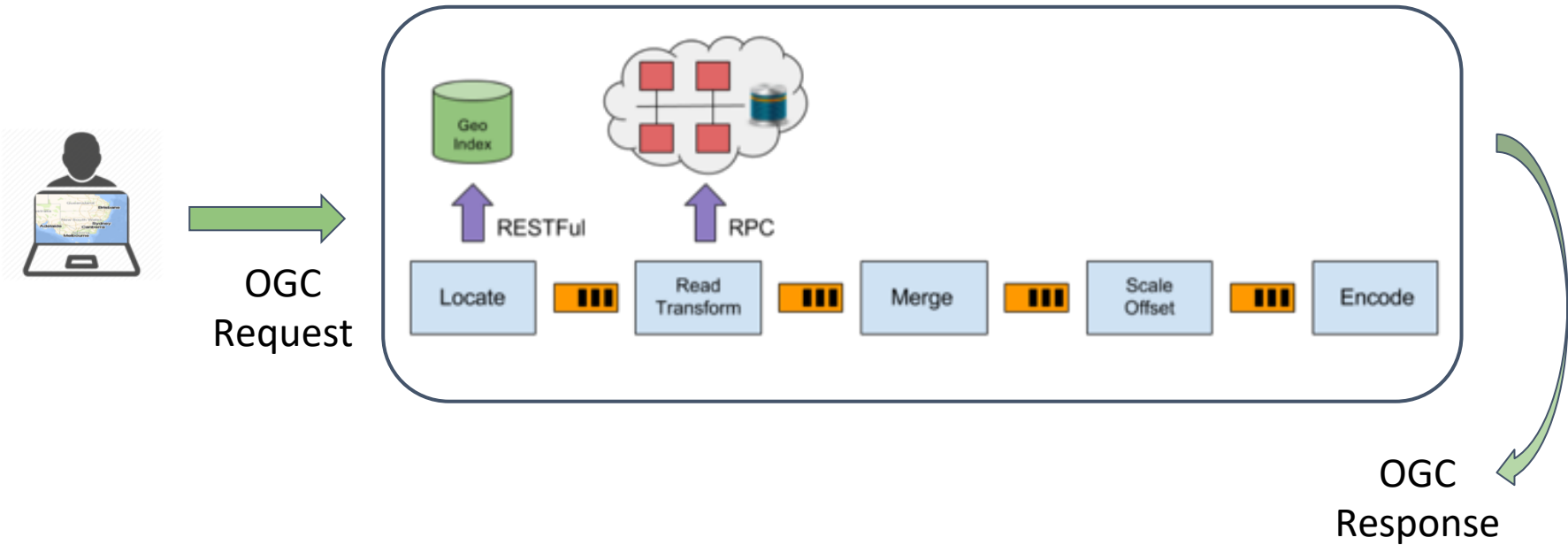
*MAS is able to process queries in milliseconds, even for the ones comprising large spatial areas or temporal ranges, which often results in thousands of data objects being identified.*

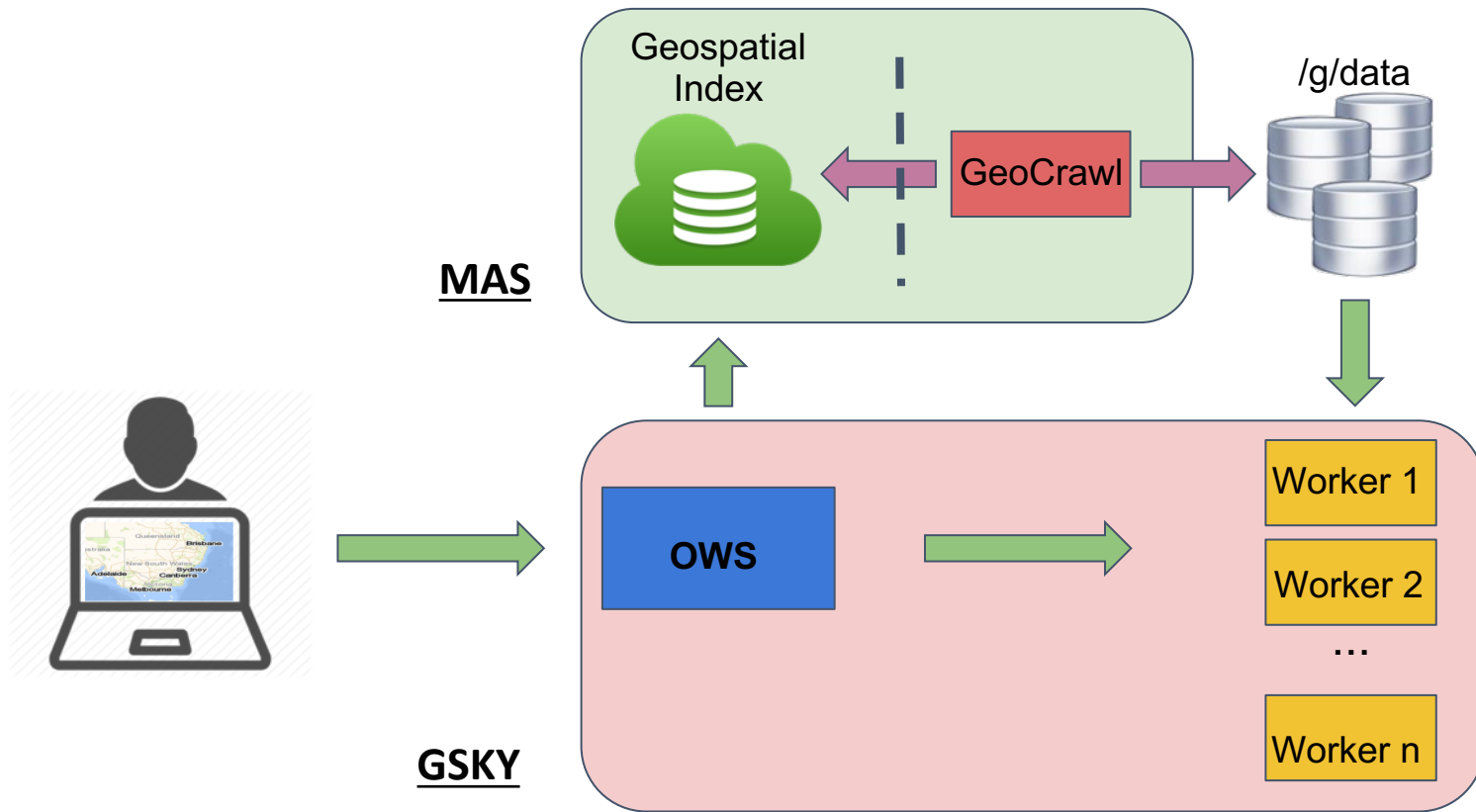
MAS is also used by other projects (e.g., our climate data services)



- Based on “flow-based programming”.
- Data transformed by connected processes, forming a Directed Acyclic Graph (DAG).

GSKY + MAS + Data + ...





## **GSKY file handling of common large Earth Observation datasets**

Amount of data is scaling up with more modern satellite instruments

### **MODIS (CSIRO & TERN)**

MODIS data from 2001 – present, with 4 timestamps per month and 1k files per timestamp.  
GSKY is handling approx. 900k MODIS files

### **Landsat-8 dataset (DEA)**

Landsat 8 dataset has data from 2013 – present, approximately 3k files per timestamp.  
GSKY is handling approx. 7 million files.

### **Sentinel 2 ARD (DEA)**

Approx 1400 available days (2015-present) and about 12k files per day for this dataset.  
GSKY is handling approx. 17 million files.

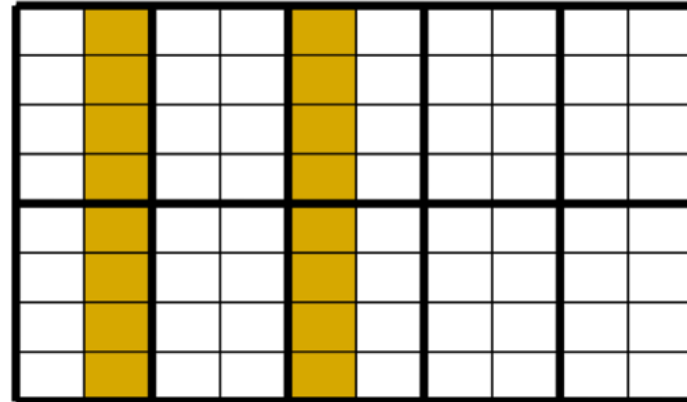
## ASTER geophysics data

Problem handling large variables

- 62 GBytes

Using:

- default TDS -> times out
- Tuned TDS -> 90s
- GSKY -> 1s



Secret: down-sampling chunked data I/O

**Digital Earth Australia (DEA) Geoscience Earth Observations**, which include the following products of the Landsat 5, 7 and 8 satellite missions:

- Surface reflectance (NBAR/NBART true and false colour)
- Terrain corrected surface reflectance geometric median (geomedian)
- Intertidal Extents Models (ITEM)
- High and Low Tide Composites (HLTC)
- Water Observations from Space (WOfS)
- Sentinel 2 Analysis Ready Data (Beta)
- Blended service (landsat + sentinel)

**GEOGLAM**, the GEO Global Agricultural Monitoring initiative, which include the following products:

- MODIS Total Vegetation Cover v3.1 (8-day and Monthly)
- MODIS Total Vegetation Cover Anomaly v3.1 (Monthly)
- MODIS Total Vegetation Cover Decile v3.1 (Monthly)
- MODIS Vegetation Fractional Cover 8-day v3.1 (8-day and Monthly)
- CHIRPS Precipitation v2.0 (Monthly)

You can browse and search NCI's full collection by going to the [Terria Map](#) or [National Map](#) websites.

To view the DEA or GEOGLAM collections, click on Add Data -> My Data -> Add Web Data and enter the following URLs respectively:

<http://gsky.nci.org.au/ows/dea>

<http://gsky.nci.org.au/ows/geoglam>

We are continually adding new datasets to GSKY.



Intro to NCI's managed FAIR datasets

Finding Data and accessing it

Intro to NCI's data services

**Data communication channels**

## Info on how we will keep them informed about data

- Webinars
- NCI monthly newsletter
- Communication to users connected to datasets or NCI accounts (depending on the info to be conveyed)
- Opus space (under construction...)

Training events:

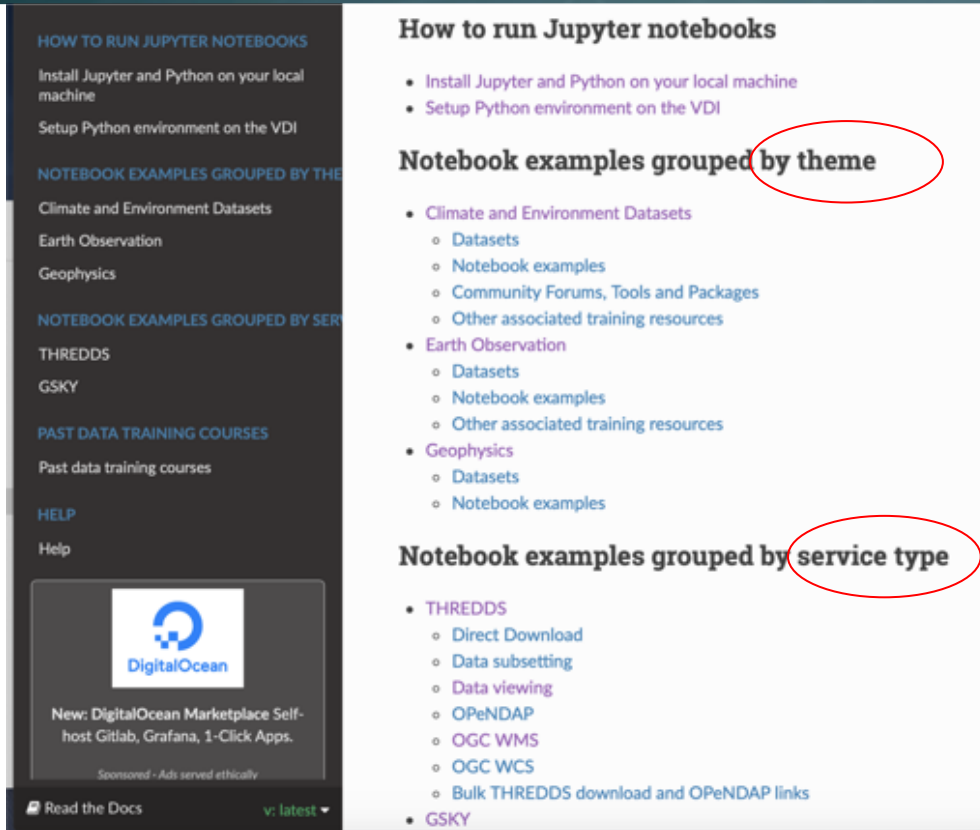
- <https://nci.org.au/news-events/events>

Data Training examples (work in progress):

- <https://nci-data-training.readthedocs.io/en/latest/>

NCI user guide (work in progress):

- <https://opus.nci.org.au>



**HOW TO RUN JUPYTER NOTEBOOKS**

Install Jupyter and Python on your local machine

Setup Python environment on the VDI

**NOTEBOOK EXAMPLES GROUPED BY THEME**

Climate and Environment Datasets

Earth Observation

Geophysics

**NOTEBOOK EXAMPLES GROUPED BY SERVICE TYPE**

THREDDS


GSKY

**PAST DATA TRAINING COURSES**

Past data training courses

**HELP**

Help



New: DigitalOcean Marketplace Self-host Gitlab, Grafana, 1-Click Apps.

Sponsored - Ads served ethically

[Read the Docs](#) v: latest

### How to run Jupyter notebooks

- Install Jupyter and Python on your local machine
- Setup Python environment on the VDI

### Notebook examples grouped by theme

- Climate and Environment Datasets
  - Datasets
  - Notebook examples
  - Community Forums, Tools and Packages
  - Other associated training resources
- Earth Observation
  - Datasets
  - Notebook examples
  - Other associated training resources
- Geophysics
  - Datasets
  - Notebook examples

### Notebook examples grouped by service type

- THREDDS
  - Direct Download
  - Data subsetting
  - Data viewing
  - OPeNDAP
  - OGC WMS
  - OGC WCS
  - Bulk THREDDS download and OPeNDAP links
- GSKY

- Time : fortnightly on Tuesday at 11:00am (Canberra Time)
- Info: <https://opus.nci.org.au/display/Help/NCI+Data+Webinar>
- Format: lecture/live demo + Q&A
  
- Topics will cover (but not limit to):
  - Walk-through of various [data services](#) at NCI, including how to find datasets and access them
  - Information on specific datasets, including a chance to talk to key data owners and users
  - Information on NCI [dataset management](#)
  - Examples of how to use the datasets using popular applications, portals, software, [Virtual Research Environments](#)
  - Other relevant use-cases for specific communities