



ESGF within NCI

- NCI part of Government's broad National Research Infrastructure Review, which includes nearly all large scale national infrastructure
- National "mood" has largely moved from century scale to seasonal-to-decadal
- CMIP activity seen as important national role for ACCESS/Climate research but pressure for more practical outcomes. e.g. food security, agriculture, water mgt, environmental, urban impacts.
- CMIP part of a number of significant data resources that are important to make interoperable with HPC/HPDA.
- Value of NCI acknowledged but still large number of questions about funding futures. e.g. who pays for storage
- NCI has a cloud but should we have more in commercial clouds and have users pay and is it beneficial to us?

- NCI has top 100 supercomputer, a high performance OpenStack cloud, large Lustre site-wide storage for datasets, and deep tape archive.
- NCI other key environmental, earth system and solid earth simulation, data and data analysis. CMIP is currently ~10% of the total volume of interest. i.e. 10% of data under ESGF type management.
- Has developed processes for managing all types of simulation and observation data inc point and line data, licensed and some auth.
- Datasets include (<http://datacatalogue.nci.org.au.au>)
 - 1. Climate/ESS Model Assets and Data Products
 - 2. Earth and Marine Observations and Data Products
 - 3. Geoscience Collections
 - 4. Terrestrial Ecosystems Collections

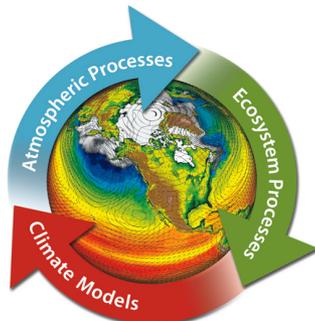
- Major model to be published from Australian will be the ACCESS-CM2 model and possibly an ESM model.
- Additional MIPs and CORDEX data to be published, but expect smaller and perhaps more obs4MIP
- Expect replicated data to be ~5 Pbytes
- Master node of the Australian datasets (including CMIP5 era)
- full replica of the international index
- ongoing replication of the “key variables”, and republishing
- Expect more systematic assignment of DOIs
- Data replication with major ESGF nodes via gridftp end-points

- Most serious users will access via direct access to NCI filesystems using either supercomputer, VDI access or virtual labs
- Data will be indexed via our standard (non-ESGF) data catalogues – including geospatial index as well as catalogue systems (geonetwork + data discovery portal)
- Publishing data via non-authenticated OPeNDAP and OGC services
 - Expect more via authenticated methods in future
- Climate expect more data portal services (e.g., climate4Impact and other international examples available locally)
- Expect more WPS service (e.g. birdhouse, zoo, bespoke)

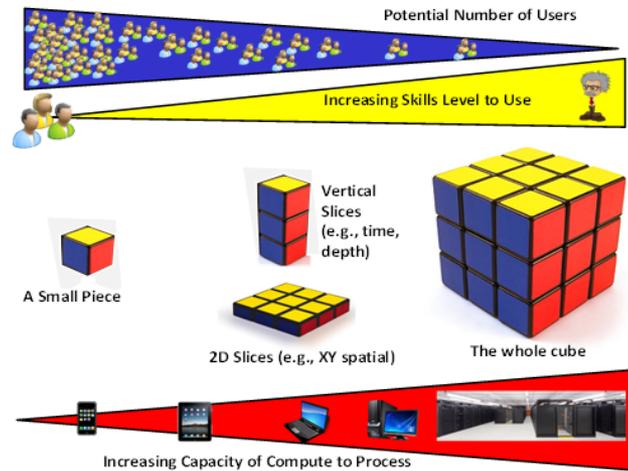
- In-situ and data service approach to data analysis
 - Need to move away from “data download/shopping cart”
 - Non-bespoke authentication/auth approach to access data
- ESGF/UV-CDAT tools available within a broader VDI service that allows interactive login.
- Broad range of data analysis, workflow and visualisation requirements that goes across very broad science domains
- Data available in-situ and via services for other workflows and publication
- Data workflows with hooks for reproducibility, PROV standard capture and publishing

- Puppet-delivered deployments in VMs using full repeatable processes for redeploy. RPMs preferred.
- Old-style bash script approach and partial upgrade to hack solutions are not sustainable.
- Docker not supported (as yet) because of security concerns.

1. Robust publishing and well-known procedures for making data available at nodes and how available internationally
2. Full documentation about datasets available.
3. Advertising international dataset availability and changes
4. Automated and fast data replication and updates for node-subscribed datasets/variables onto the local node and to remote
 1. Must be tested and work with data publication and data mgt.
5. Bullet-proof software and data update processes (i.e., test environment first)
6. Ensure that all other compute and viz software developments are available across the nodes. (I see as a major gap at the moment).



- NWP and Forecasts
UM, APS3 (Global, Regional, City), ACCESS-TC
- Coupled Seasonal and Decadal Climate
ACCESS-GC2/3 (GloSea5)
- Data Assimilation
3D-VAR, 4D-VAR (Atmosphere), EnKF (Ocean)



- Water availability and usage over time
- Catchment zone
- Vegetation changes
- Data fusion with point-clouds and local or other measurements
- Statistical techniques on key variables



Ocean Forecasting and Research
OceanMaps, BlueLink, MOM5,
CICE/SIS, WW3,
ROMS

Compute Intensive



Virtual Laboratories



Portal views



Machine Connected



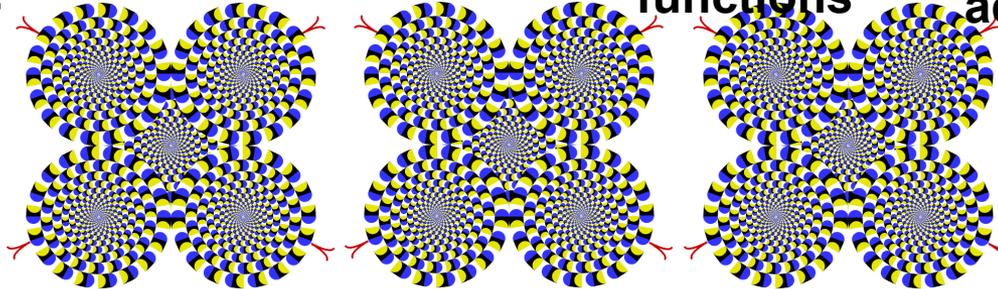
NERDIP Data Platform

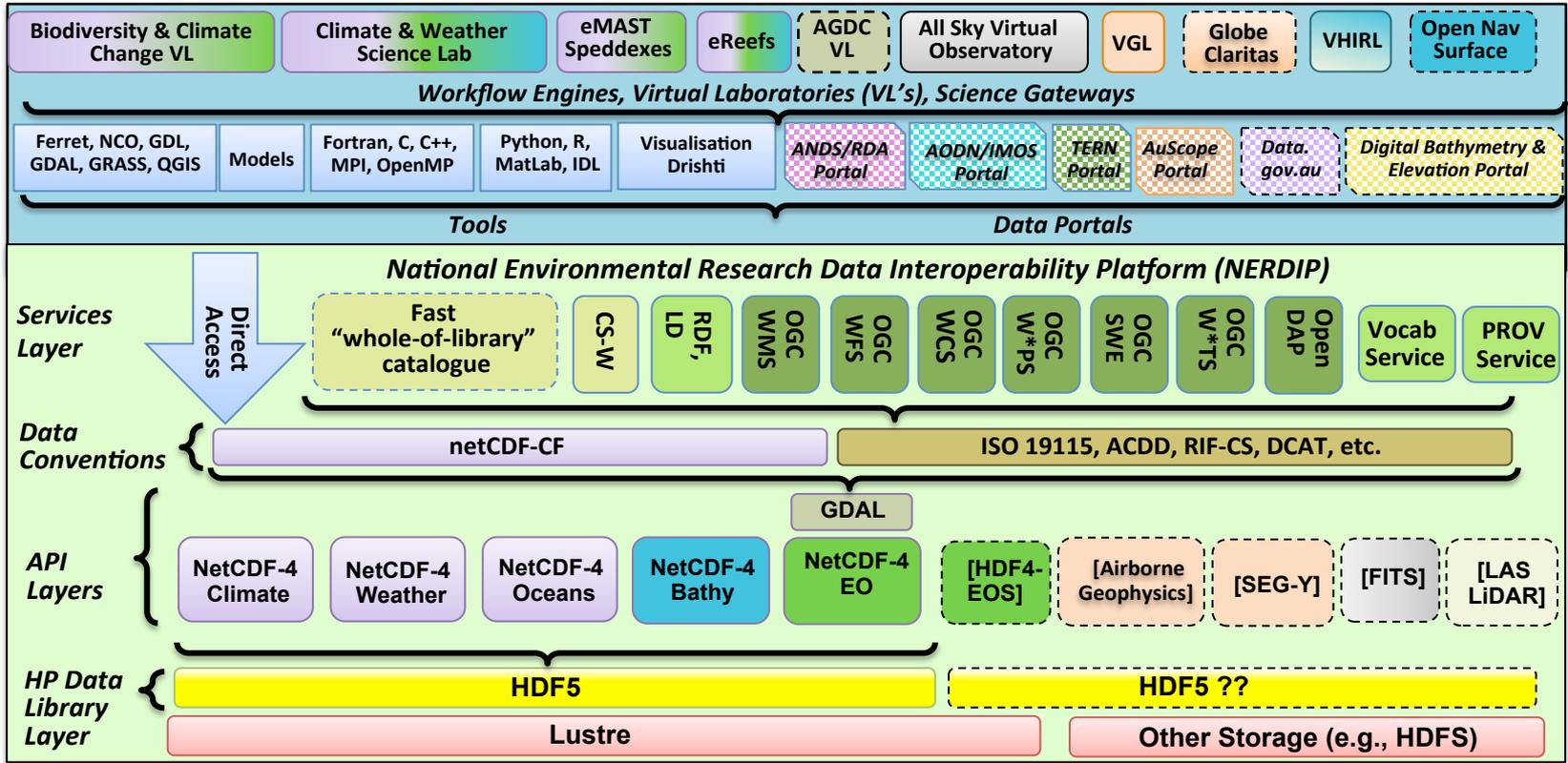
Fast/Deep Data Access

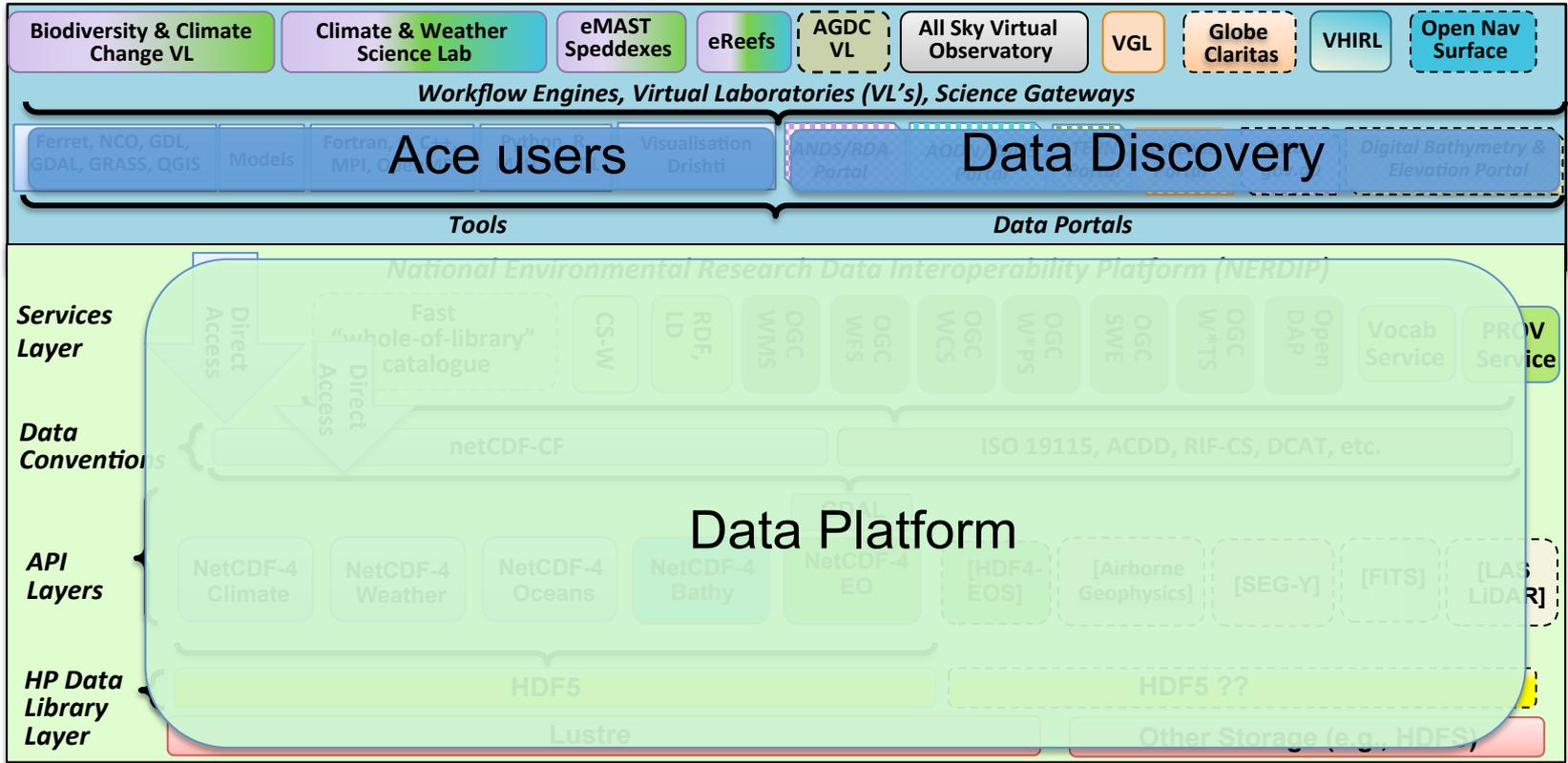
Data Services

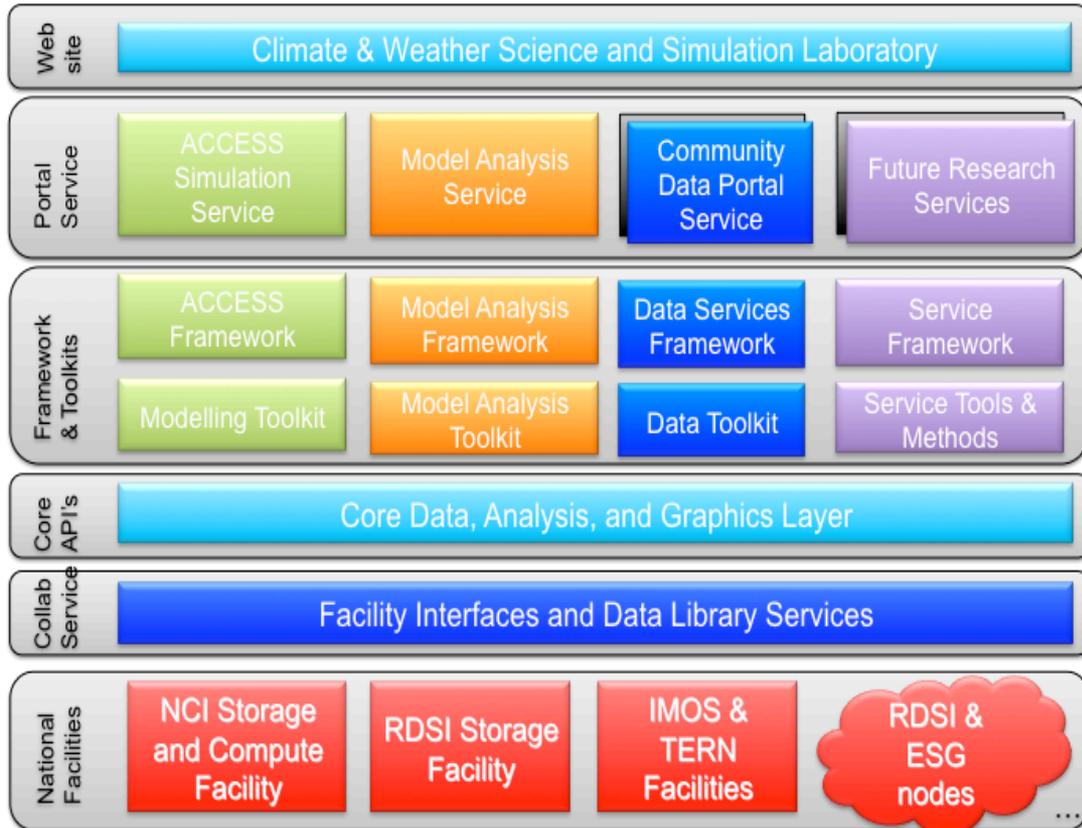
Server-side functions

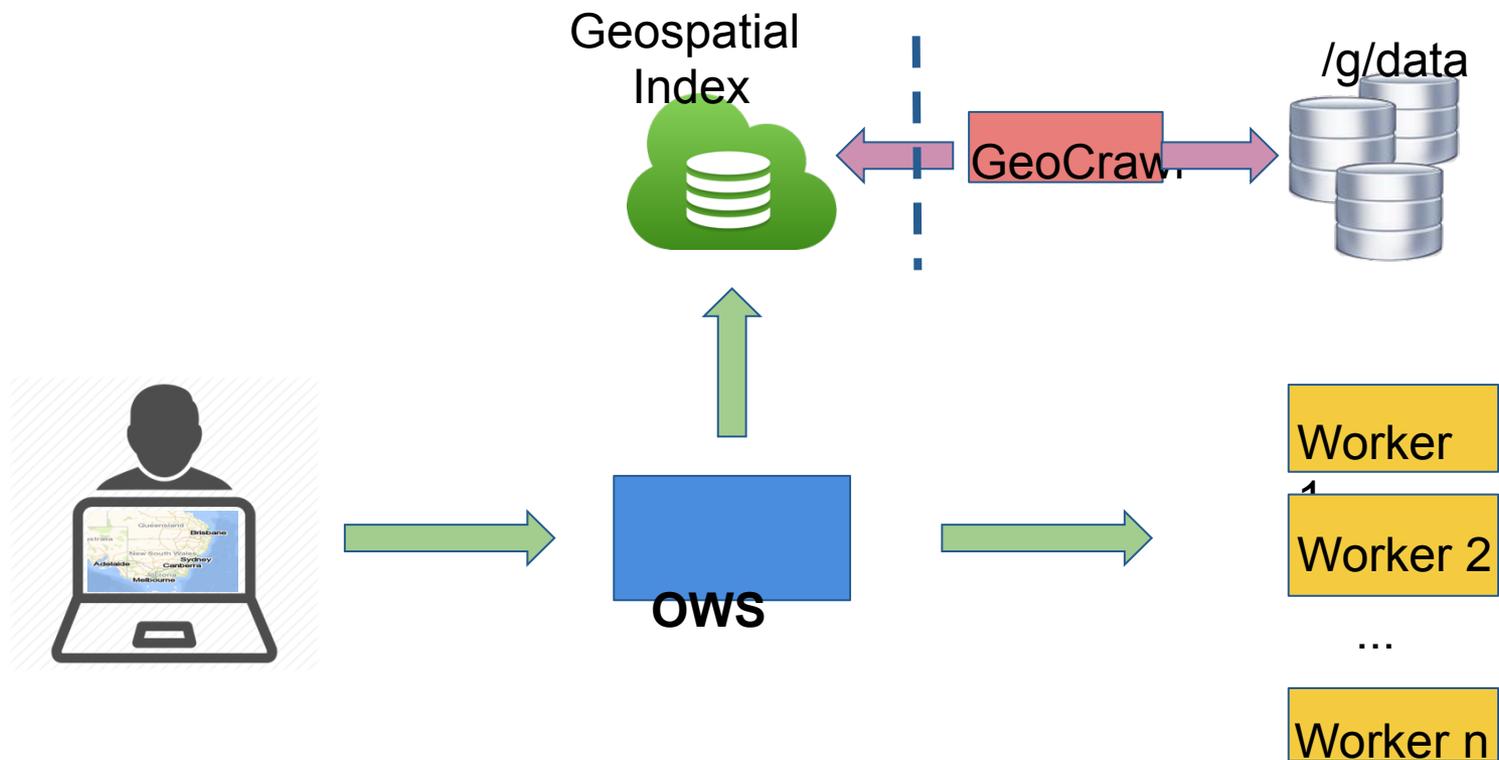
Program access











QuickTime Player File Edit View Window Help

NationalMap x

nationalmap.gov.au

NATIONAL MAP Australian Government

Search

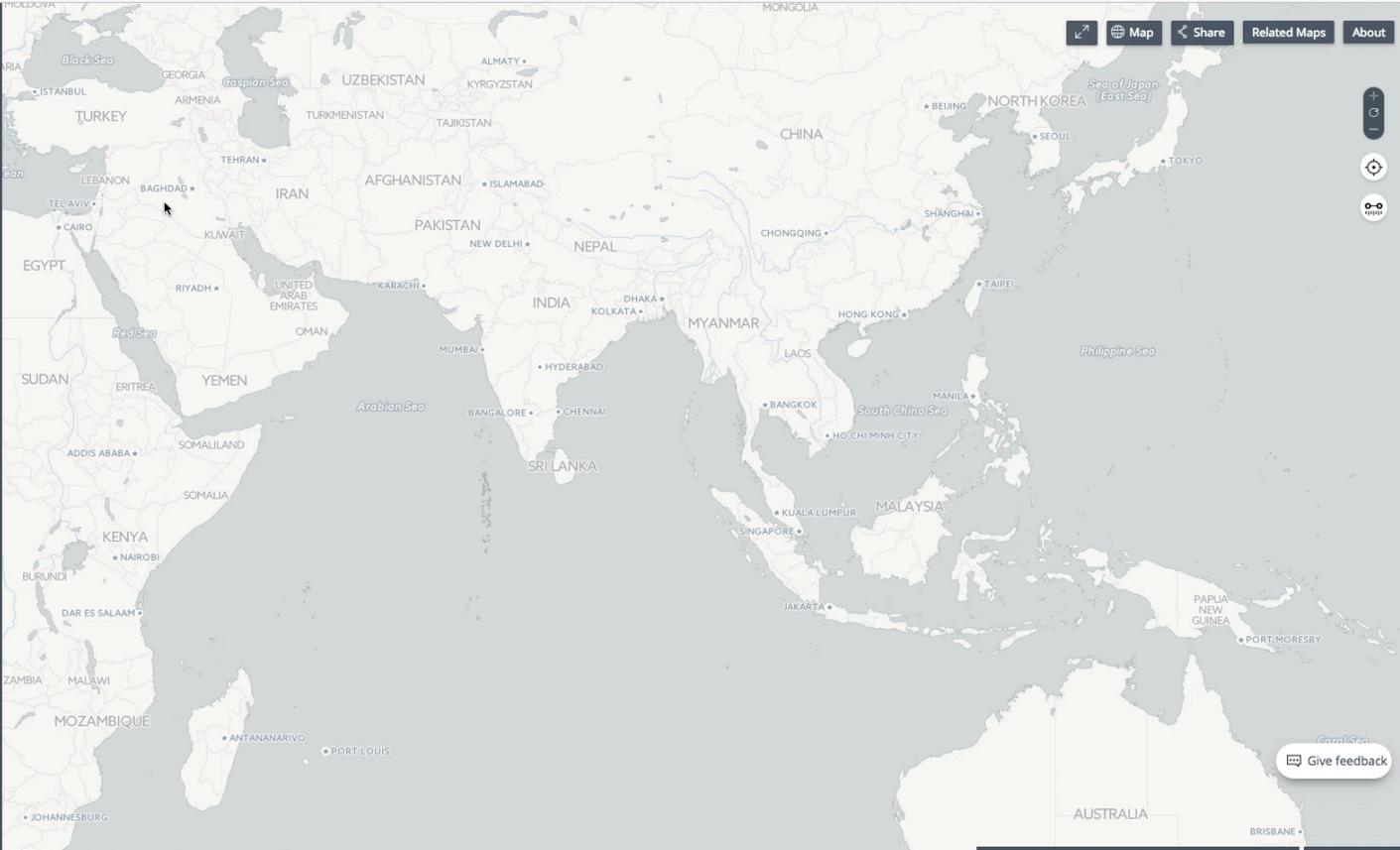
Add data

Your workbench is empty

Click 'Add Data' above to:

- Browse the Data Catalogue
- Load your own data onto the map

TIP: All your active data sets will be listed here

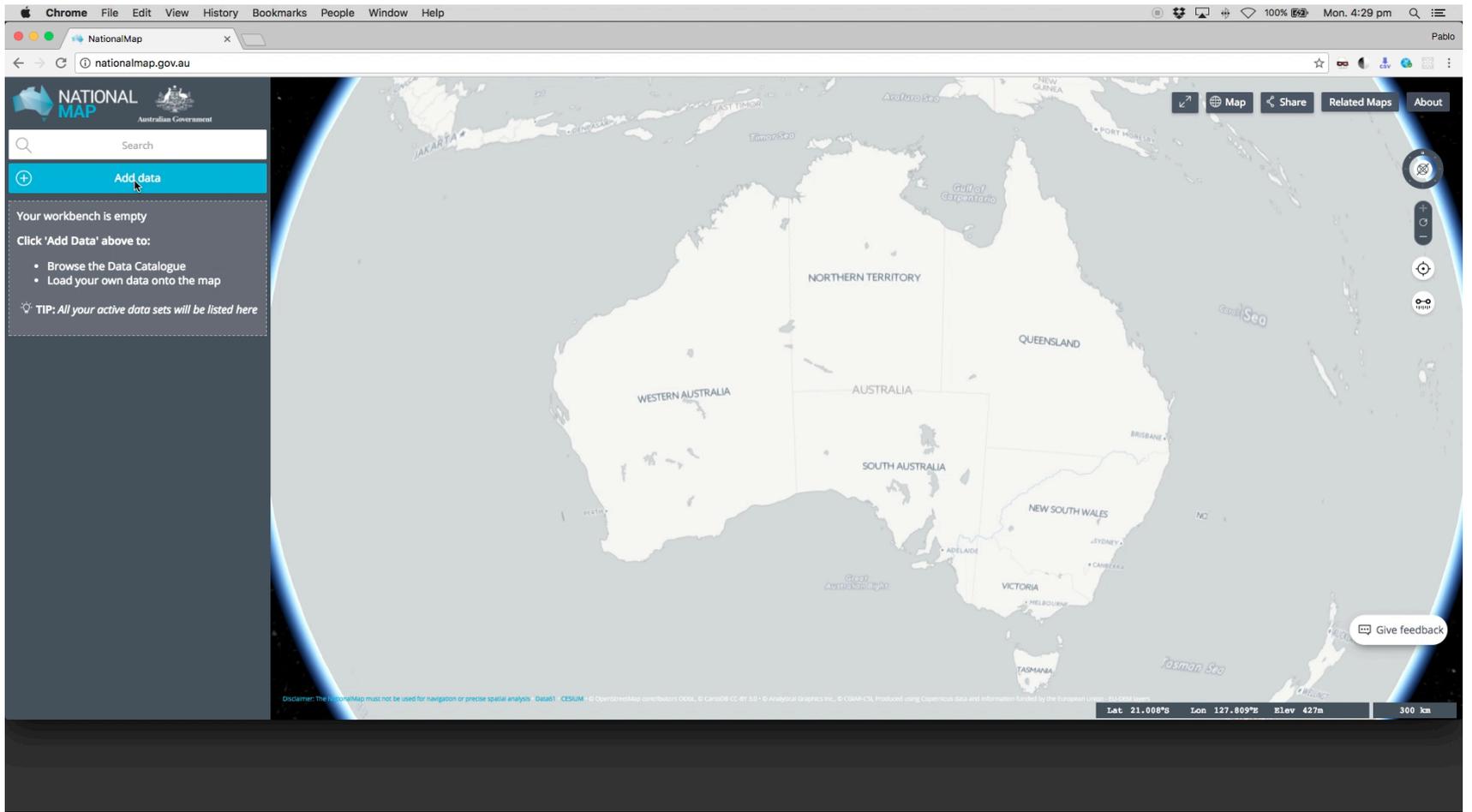


Map Share Related Maps About

Give feedback

Lat: 23.725°N Lon: 53.965°E 11m 500 km

Disclaimer: The NationalMap must not be used for navigation or precise spatial analysis. Data ©1 Leaflet | © OpenStreetMap contributors ODbL, © CartoDB CC-BY 3.0



The screenshot shows the NationalMap.gov.au web application in a Chrome browser window. The browser's address bar displays 'nationalmap.gov.au'. The application interface includes a top navigation bar with 'Map', 'Share', 'Related Maps', and 'About' buttons. A search bar is located in the top left, with an 'Add data' button below it. A sidebar on the left contains the following text:

NATIONAL MAP
Australian Government

Search

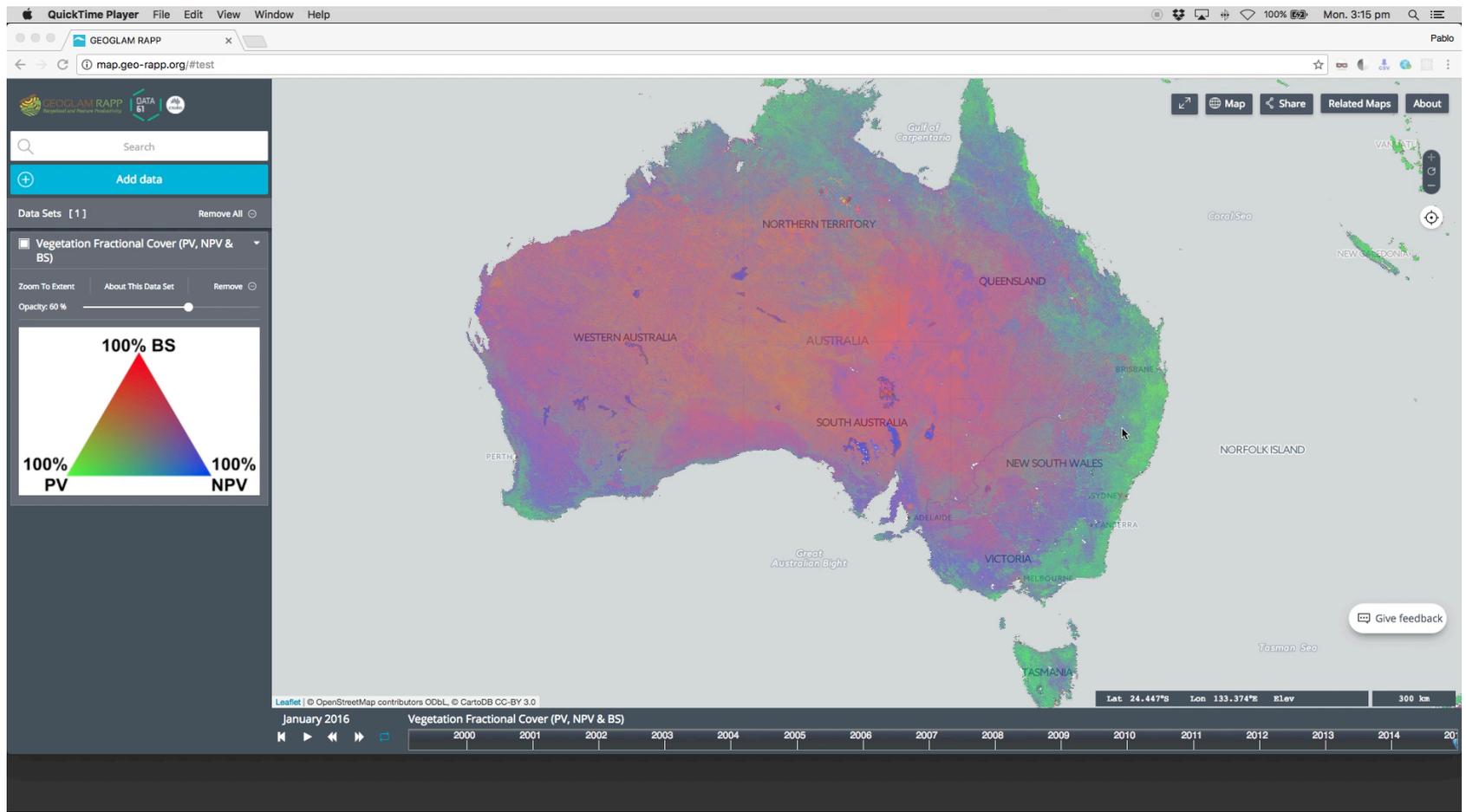
Add data

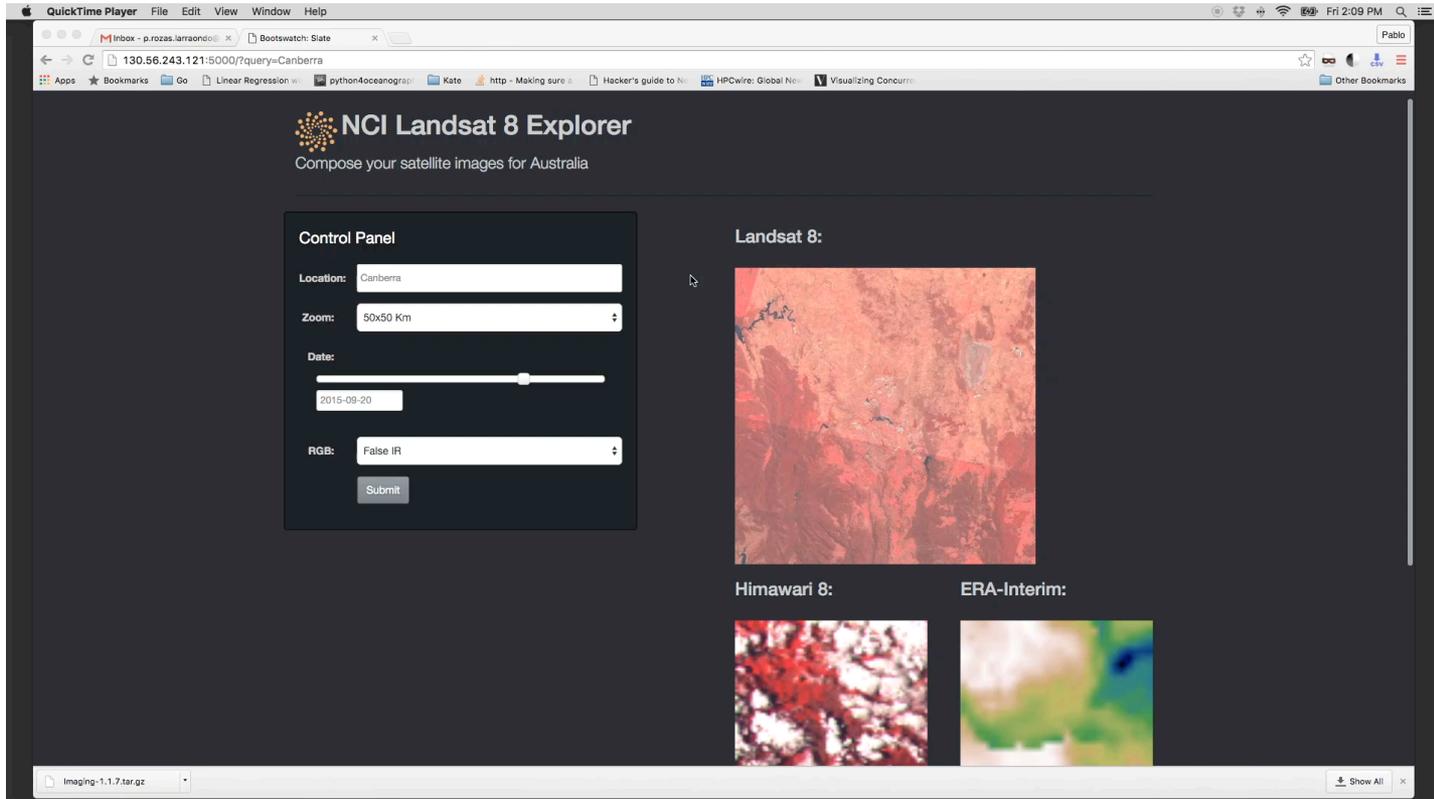
Your workbench is empty
Click 'Add Data' above to:

- Browse the Data Catalogue
- Load your own data onto the map

TIP: All your active data sets will be listed here

The main map area displays a map of Australia with state and territory labels: WESTERN AUSTRALIA, NORTHERN TERRITORY, AUSTRALIA, QUEENSLAND, SOUTH AUSTRALIA, NEW SOUTH WALES, VICTORIA, and TASMANIA. Major cities like PERTH, ADELAIDE, SYDNEY, MELBOURNE, and BRISBANE are also labeled. The map is surrounded by a blue globe border. A 'Give feedback' button is visible in the bottom right corner of the map area. At the bottom of the application, a status bar shows coordinates: Lat 21.00678, Lon 127.80978, Elev 427m, and a scale of 300 km.



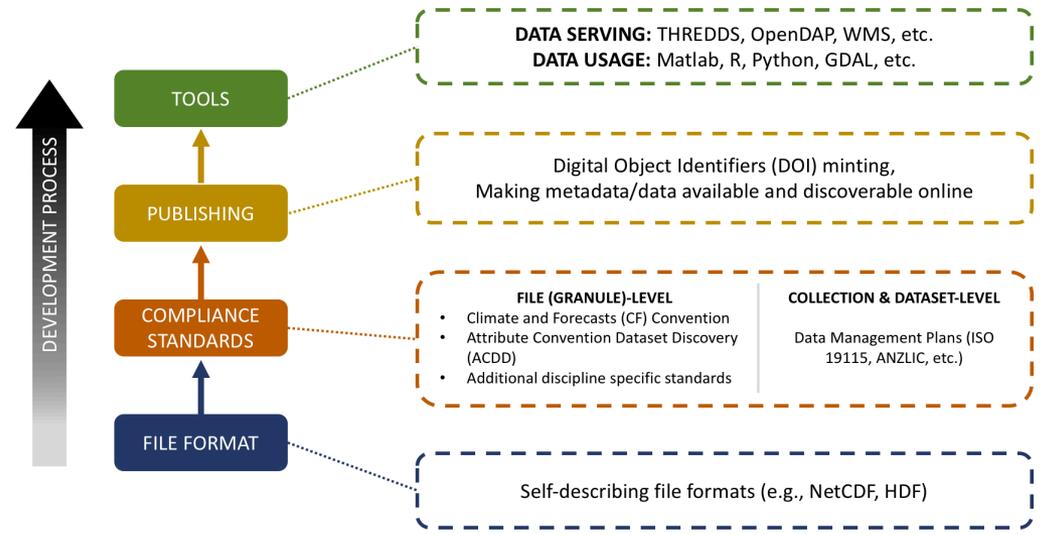


The screenshot shows a web browser window displaying the "NCI Landsat 8 Explorer" application. The browser's address bar shows the URL "130.56.243.121:5000/?query=Canberra". The application interface includes a "Control Panel" on the left with the following settings: Location: Canberra, Zoom: 50x50 Km, Date: 2015-09-20 (with a slider), and RGB: False IR. A "Submit" button is located below these controls. On the right, there are three satellite image thumbnails: "Landsat 8:" (a large red-toned image), "Himawari 8:" (a smaller red-toned image), and "ERA-Interim:" (a smaller green and blue image). The browser's status bar at the bottom shows "Imaging-1.1.7.tar.gz" and a "Show All" button.



Data Quality Strategy (DQS): What does it involve?

1. Underlying HPD file format
2. Close collaboration with data custodians and managers
 - Planning, designing, or reassessing the data collections



3. Quality control through compliance with recognised community standards



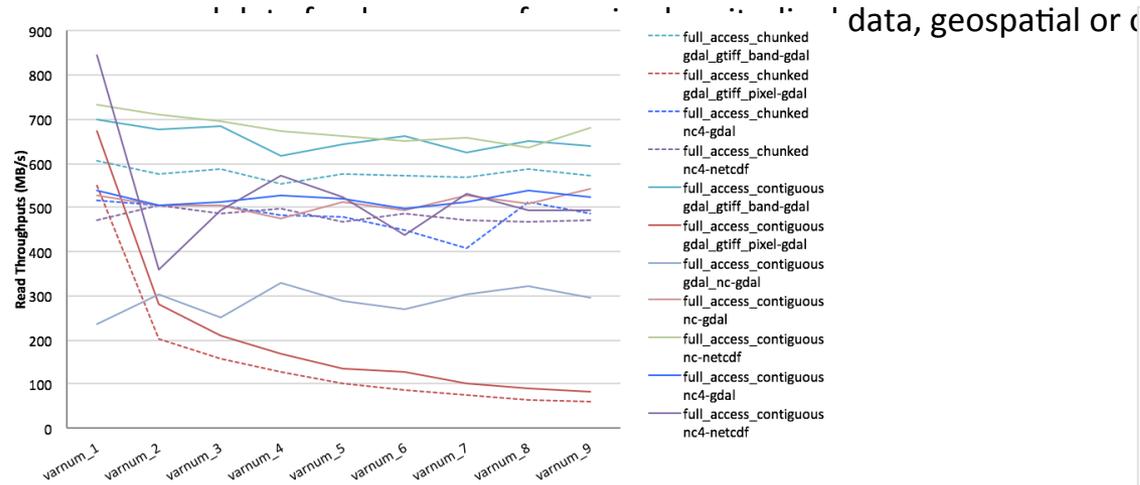
4. Data assurance through

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	

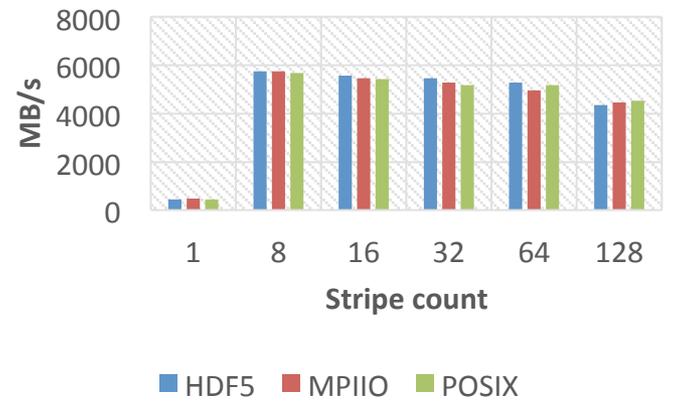


- Quality Assurance against performance metrics
- We need to scale data so that you can analyse in real-time and in-situ.
- Need to combine/overlay/slice-dice all manner of data at high precision from vast reference with highly specific data.
- We need faster, automated systems for real world activities, decision making capability using smart new algorithms and programmatic techniques:



data, geospatial or c

Independent Read



NCI AUSTRALIA

WOFST PIXEL DATA

1988 1990 1992 1994 1996 1998 2000

Godiva

Layer: NCI THREDDS Server - OBR4 Hydro - Temperature
Units: degrees C
Height (m): 0.5
Date/Time: 31 Oct 2015 00:00:00 UTC Spot Range: last 30 days

Day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
1	1	2	3				
4	5	6	7	8	9	10	
11	12	13	14	15	16	17	
18	19	20	21	22	23	24	
25	26	27	28	29	30	31	

ARCGIS

SPEDDEXES

Search: temperature

Default: The distribution pattern of monthly daily maximum temperature (you can select from a list of 100000)

Variables: Monthly daily maximum temp

Available Processes: GetFullSeries, Plot timeseries

ESGF Earth System Grid Federation

Transect Data

14-day SST MODIS mosaic (lat, mosaic)

13 Mar 2016 01:20

World Wind Explorer

Layers: Bing Aerial with Labels, Open Street Map, NCI THREDDS Server

EarthServer rasdaman raster data management



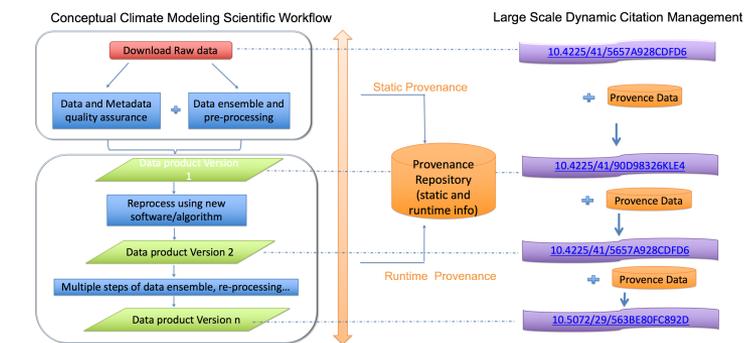
- PROMS v3 uses an extension to the PROV ontology as its data model.
 - Entities
 - Activities
 - Agent

NCI PROMs Server
Provenance Management Service

An Activity
URI: <http://130.56.244.85#10w3871-74c4-4438-8666-26f1832597>

Title: CC Whole Process
Started at: 2014-06-25T12:13:14
Ended at: 2014-06-25T12:13:14
Was associated with: <http://130.56.244.85#1c46684-9947-4b01-8c3f-61377499738>
Report: <http://nci.australia.gov.au/2014-06-25>

Neighbours view



The Provenance Service captures information at each step within the end-to-end workflow, and stores it within the Provenance Repository

RD-Switchboard <http://www.rd-switchboard.org/>

