

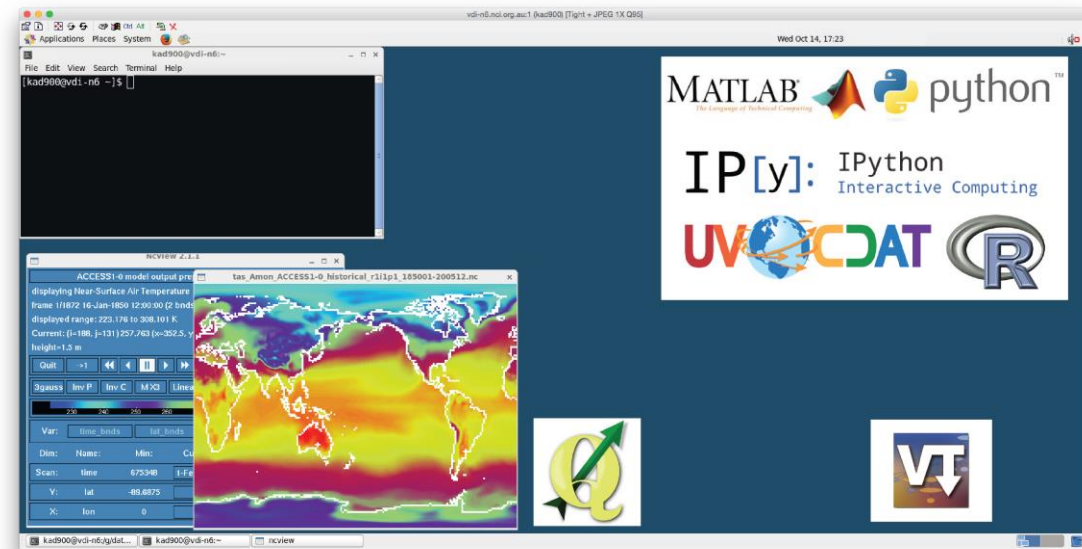


NCI
AUSTRALIA

The Australian Climate Science Data-enhanced Virtual Laboratory

*Clare Richards, Ben Evans, Kate Snow, Chris Allen, Matt Nethery,
Paola Petrelli, Claire Trenham, Aurel Moise, Sean Pringle,
Claire Carouge, Scott Wales, Louise Wilson, Tim Erwin.*

- The history
- The Climate Science DeVL project
- Preparing for the future



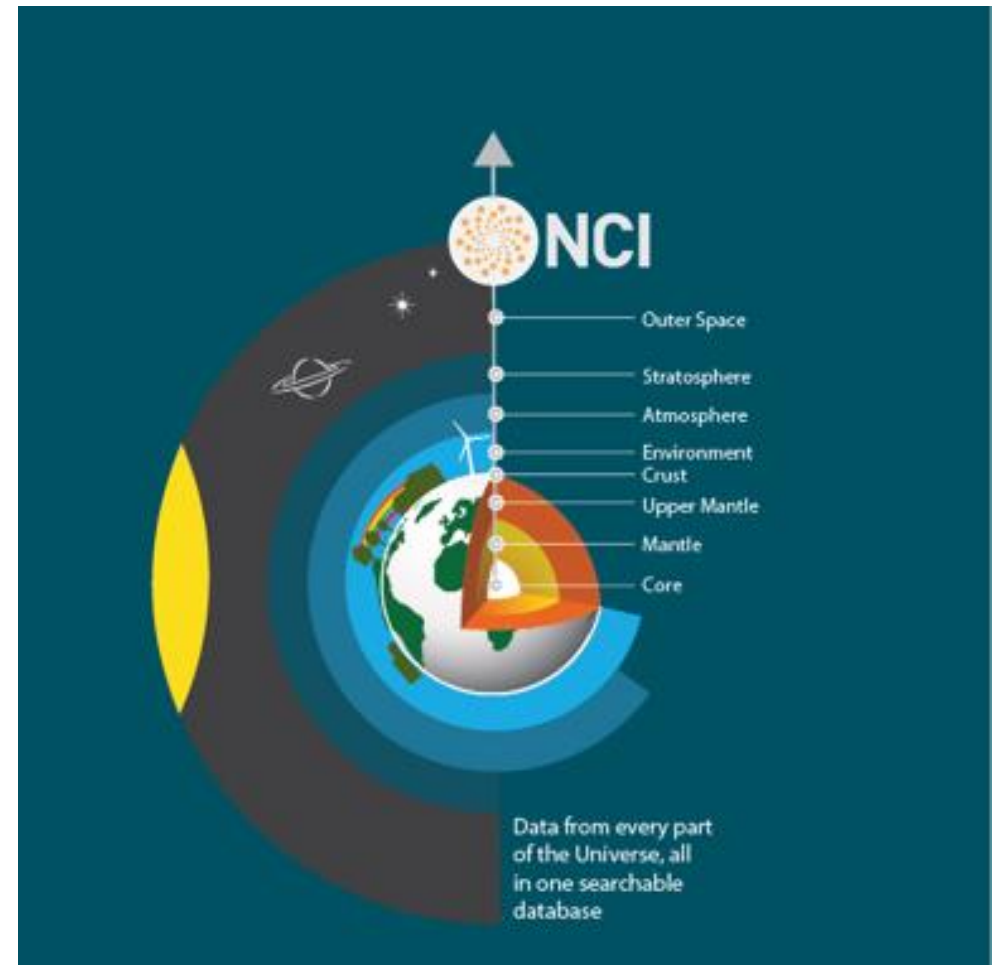
HPC and cloud infrastructure

- Computationally- and data-intensive science
- Big domains and long simulations

15+PB of reference data collections:

- **Climate and Weather**
- Environmental
- Earth Observation
- Geophysical
- Optical Astronomy

- Other data:
Genomics and Social Sciences



One of the most **computationally demanding** research domains in the environmental sciences.

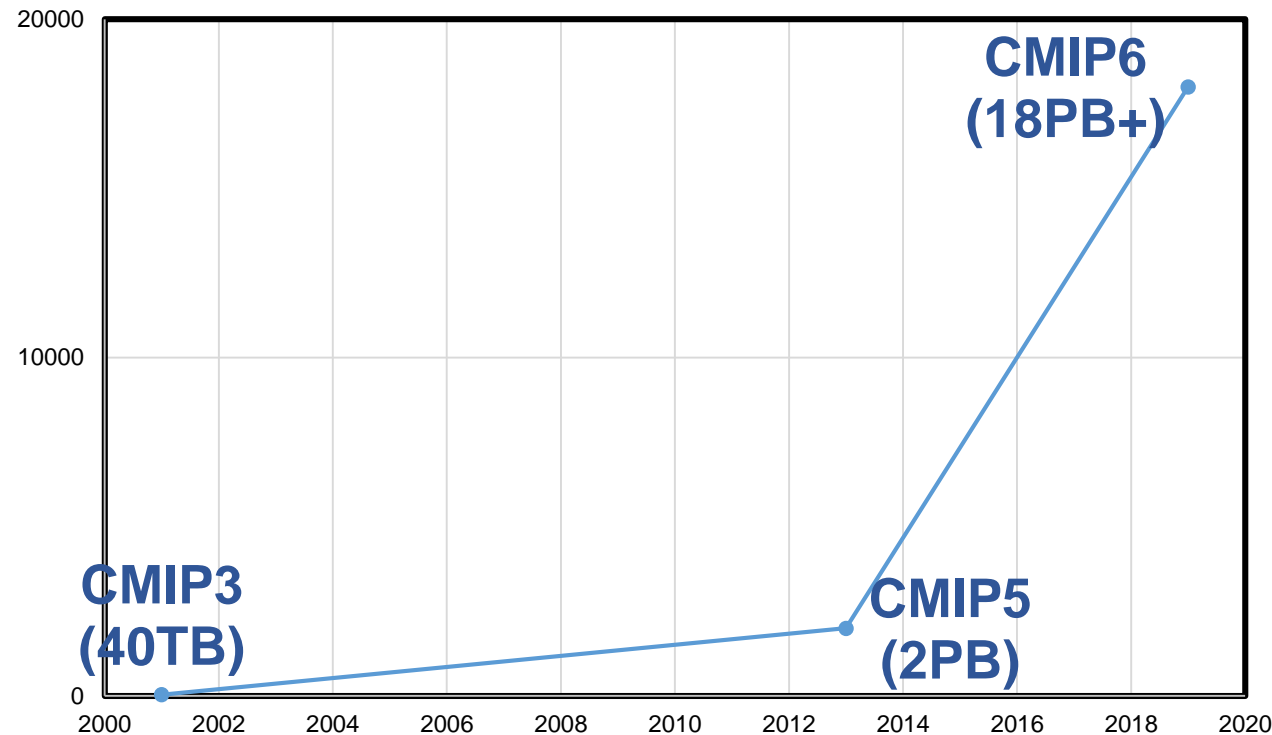
Volume and complexity of data is growing

- International collaboration involves sharing a lot of data.
- Coupled Model Intercomparison Project (CMIP)

Need a research platform for model development and intensive data analysis

- Must adapt as user needs change.

Growth in CMIP Data



Balaji V, et al (2018) *Requirements for a global data infrastructure in support of CMIP6*

Climate and Weather Science Virtual Laboratory (2013-2016)

- Integrated compute and data analysis platforms

Research Data Storage and Services (2011-2017)

- Repository for National Reference Data Collections established
- Easier access to quality data
- Co-located with Petascale HPC and Cloud

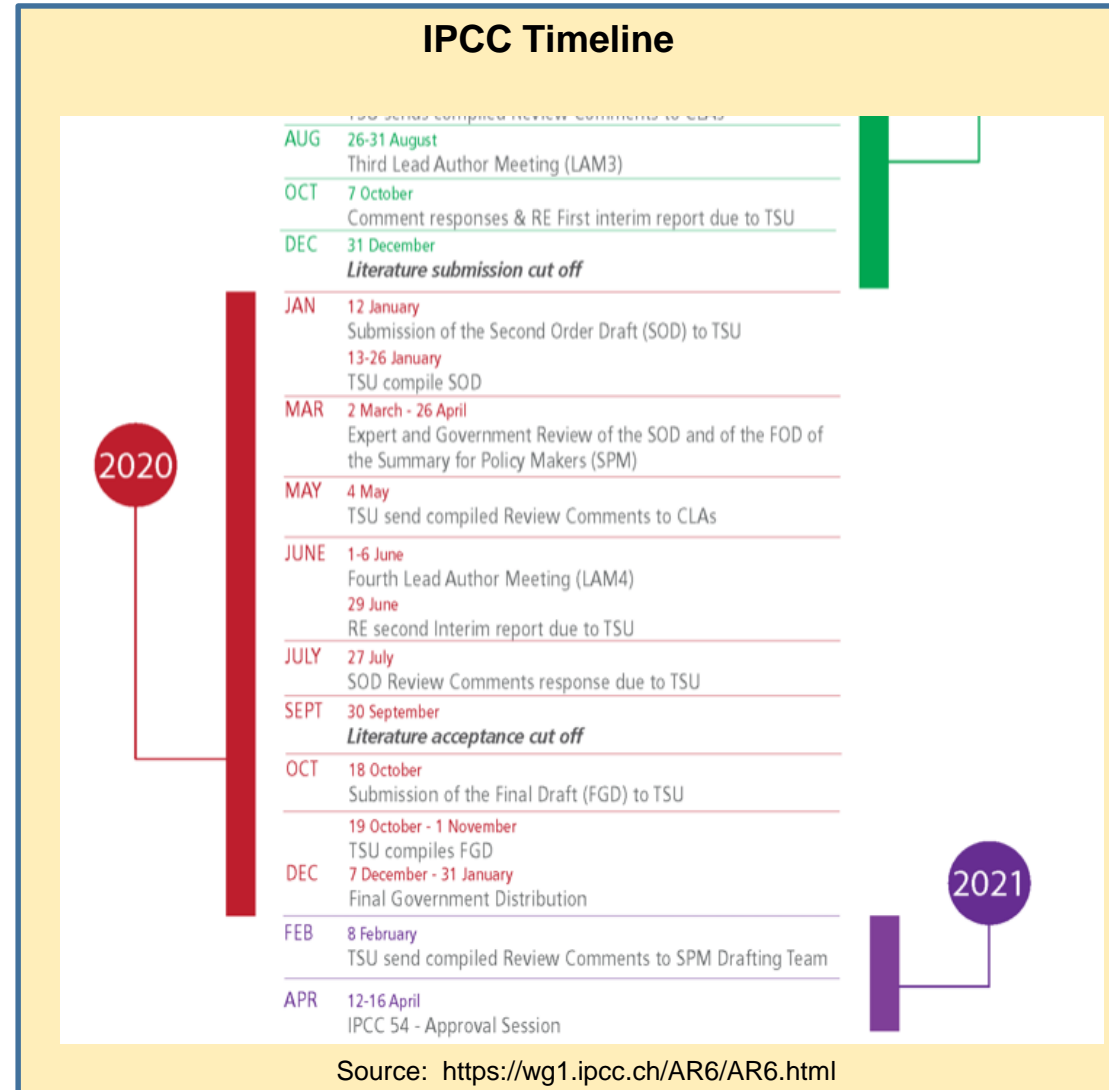
Collaborative user-focused development

- Users with range of skills, experience.



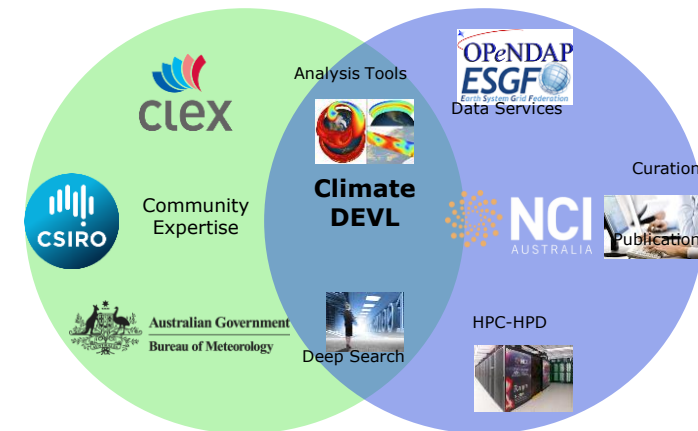
- Coupled Model Intercomparison Project Phase 6 (CMIP6)
 - World Climate Research Programme (WCRP)
 - Internationally coordinated research activity
 - Basis of assessments by the IPCC (Intergovernmental Panel on Climate Change)
 - Informs global climate agreements
 - Used by Australian research, government, business, agriculture and industry

- Australian Research effort for CMIP6
 - Model Development (~~ACCESS-CM2, ACCESS-ESM1.5~~)
 - Data management, access and analysis



Climate Science Data-enhanced Virtual Laboratory (DeVL) project (2018-19)

- Reliable access to Climate and Weather data collections (local and internationally)
 - CMIP6 will be biggest climate dataset ever!
 - Earth System Grid Federation (ESGF) software upgrades
- Updated data analysis environments at NCI
 - Improved tools
- Improved FAIRness of data
 - Scalable data search
- User support and training



ESGF Deployment and CMIP data replication and management

- a. International coordination - NCI Tier-1 site
- b. New ESGF node deployed (<https://esgf.nci.org.au>)
 - i. Components of stack are separate

- a. Replication and management of international data
 - i. Synda software deployed on DTN
 - ii. Fully evaluated by redownload of CMIP5 (~615TB)
 - iii. CMIP6 download approx. 42TB but expecting 3PB+
 - iv. Majority of download mid-late 2019

- a. Publishing Australian data to the ESGF
 - i. Upwards of 400TB (mid year)

For more information on ESGF:
<https://esgf.llnl.gov/>

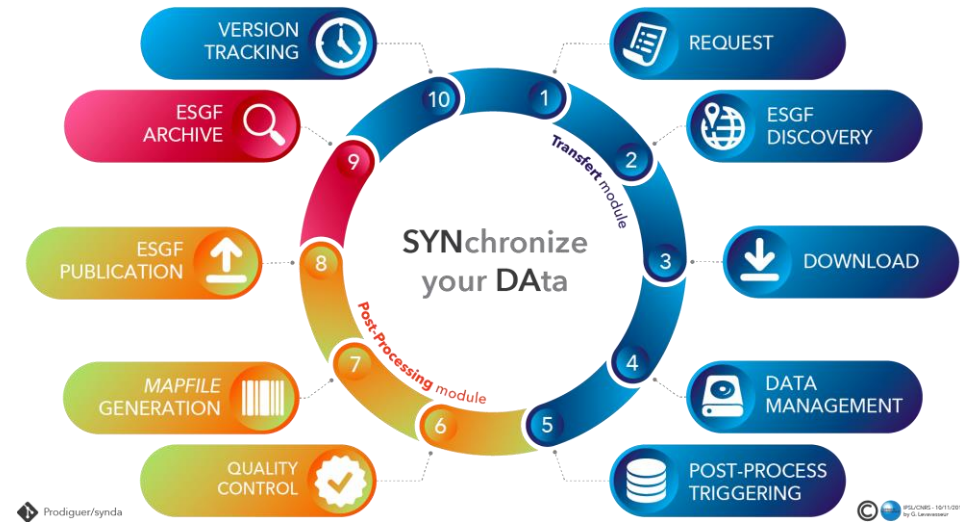
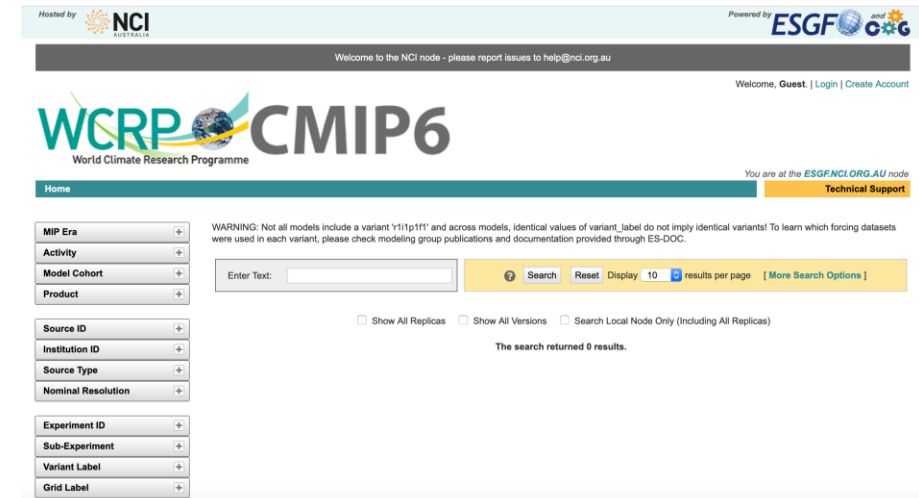



Image Source: <http://prodiguer.github.io/synda/sdp/replication.html>

- Errata Service coordination and updates
- Data Management Planning
 - CMIP6 expected at 20PB, approx. 6PB likely requirement for Australian climate community.
 - Finite storage available.
- Data Management approach
 - Increased focus on quality and reliability
- Data Publication Services
 - ESGF
 - THREDDS
 - Modelling centre support

Welcome to the DM...

Total Issues = 22. Filtered Issues = 22.

#	Institute	Title	Created	Updated	Closed	Severity	Status
1	NOAA-GFDL	piControl variable Eday/zg	2019-02-07	2019-02-07	--	Low	Resolved
2	IPSL	300 years second extension for abrupt-4xCO2	2019-01-18	2019-02-21	--	Low	Resolved
3	IPSL	200 years extension for piControl	2018-11-29	2019-02-21	--	Low	Resolved
4	IPSL	"Fixed" CMIP6 variables provided by NEMO model are ti ...	2018-11-26	2019-02-13	--	Medium	Resolved
5	NOAA-GFDL	Variable tslsi (3hr,day) has incorrect "comment" vari ...	2018-11-26	--	--	Low	New
6	IPSL	500 years extension for piControl	2018-11-23	2018-11-29	--	Low	Resolved
7	CNRM-CERFACS	Wrong realm ocnBgChem typo	2018-11-14	2018-11-16	--	Low	Resolved
8	NOAA-GFDL	Incorrect some coordinates and cell_methods in piCont ...	2018-11-08	2018-11-08	--	Medium	New
9	NOAA-GFDL	Error in variable "comment" metadata	2018-11-01	2018-11-16	--	Low	New
10	NOAA-GFDL	albiscpp erroneous data units	2018-10-29	2018-11-16	--	Low	New
11	IPSL	300 years extension for abrupt-4xCO2	2018-10-22	2018-10-22	--	Low	Resolved
12	IPSL	Irrelevant CFC in experiment other than historical	2018-10-19	2019-02-13	--	Low	Resolved
13	IPSL	Instabilities which lead to erroneous values of tas a ...	2018-10-16	2019-02-13	--	Critical	On Hold
14	IPSL	tas instabilities lead to erroneous values of tasmax	2018-10-05	2019-02-08	--	Critical	On Hold
						Critical	Resolved
						Low	Resolved
						High	Resolved
						Low	Resolved
						Low	Wont Fix
						Low	Resolved
						Low	Resolved
						Low	Resolved

Hosted by  NCI AUSTRALIA

Welcome to the NCI node - please report issues to help@nci.org.au

Welcome, Guest | Login | Create Account

NCI ESGF Node

Home Contact Us

Welcome to the NCI ESGF Node

Search & Download Data

Simple Text Search [More search options](#)

The Earth System Grid Federation (ESGF) consists of federated data centres across the globe that enable access to the largest archive of climate data world-wide. This portal allows you to find, select and download data files from the federation.


The NCI ESGF node currently hosts Australian produced data for the following collections:

- CMIP5
- CORDEX
- GeoMIP
- PMIP3

Additionally, we have locally replicated copies of data collections that have been published on other ESGF nodes.

Local Data Collections

Generally, data is publicly available for download and can be accessed without a login. However, some projects may require that you login first (you'll be prompted if this is

Powered by  ESGF and CoG

You are at the [ESGF/NCI.ORG.AU](#) node

Technical Support

Federated ESGF-CoG Nodes

- ESGF@CEDA
- ESGF@DKRZ
- ESGF@DOE/LNL
- ESGF@IPSL

Read News

CMIP6 has landed
We're pleased to announce that CMIP6 replica data is now available for download via the ...

It's spring clean time!
We've undergone a major upgrade on the NCI ESGF node, including increased infrastructure

Browse Projects

This All My Tags

Parent projects (0)

Peer projects (0)

Child projects (1)

CMIP6-NCI

Enter Tag

v0.6.3.0 © ES-DOC

x.html

- CMIP datasets registered in NCI data catalogue (geonetwork.nci.org.au)
- ARDC's Research Data Australia (RDA) data discovery portal harvests these catalogue entries
- Will be harvested into Google Dataset Search



Coupled Model Intercomparison Project phase 6 (CMIP6)

Updated: 6 months ago

This data collection contains replicated output of the Coupled Model Intercomparison Project phase 6 (CMIP6) experiments.

CMIP6 will provide a multi-model context with the main aim being to improve our understanding of past, present and future climate change. Additional goals of CMIP6 include: 1) assessing the mechanisms responsible for model differences in poorly understood climate feedbacks 2) examining climate predictability and exploring the ability of models to predict climate on decadal time scales, and, more generally, 4) determining why differently forced models produce a range of responses.

The data collection originates from the World Climate Research Programme (WCRP :<https://www.wcrp-climate.org/wgcm-cmip>) Working Group of Working Group on Model Intercomparison (WGMI) with numerous climate modeling groups from around the world participating to produce one of the largest climate datasets. CMIP6 experiments include the core DECK experiments (including amip, 1pctCO2, piControl, and abrupt4xCO2) as well as historical and endorsed MIPs: <https://www.wcrp-climate.org/modelling-wgcm-mip-catalogue/modelling-wgcm-cmip6-endorsed-mips>

The collection of climate model runs for CMIP are hosted through the Earth System Grid Federation (ESGF) of which NCI is a Tier 1 node and serves as a replica repository of the priority datasets. The replica data is located locally on the NCI filesystems under `/g/data/o110`.

For more information, please refer to the NCI CMIP website: <https://opus.nci.org.au/display/CMIP/CMIP+Home>

ongoing

Spatial extent



Related and links

CMIP6 Special Issue Articles

Open link

Temporal extent

Period
»

Provided by



Share on social s

About this resource

Keywords

- Other Chemical Sciences Environmental Chemistry
- Atmospheric Sciences
- Oceanography
- Physical Geography and Environmental Geoscience
- National Computational Infrastructure NCI

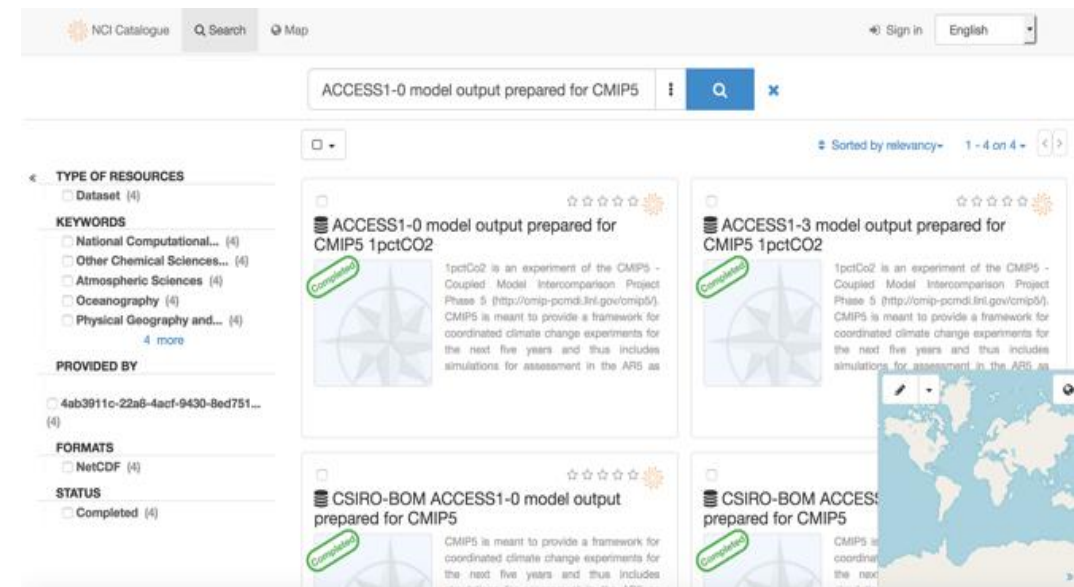
- Data index and search
 - All official CMIP data at NCI indexed by the NCI Metadata Attribute Service (MAS)
 - MAS integrated with the CleF API* for improved user search of large datasets
 - NCI ESGF node extended for overseas ESGF index
- Improvements to NCI infrastructure and services
 - Notification service for new data versions
 - Synda automated replication as new data becomes available
 - Automatic republishing to NCI ESGF node

```
[pxp581@vdi-n24 clef]$ clef --help
Usage: clef [OPTIONS] COMMAND [ARGS]...

Options:
  --remote  returns only ESGF search results
  --local   returns only local files matching ESGF search
  --missing returns only missing files matching ESGF search
  --request send NCI request to download missing files matching ESGF search
  --help    Show this message and exit.

Commands:
  cmip5 Search local database for files matching the...
  cmip6 Search local database for files matching the...
```

*See Claire Trenham's talk ... on now!

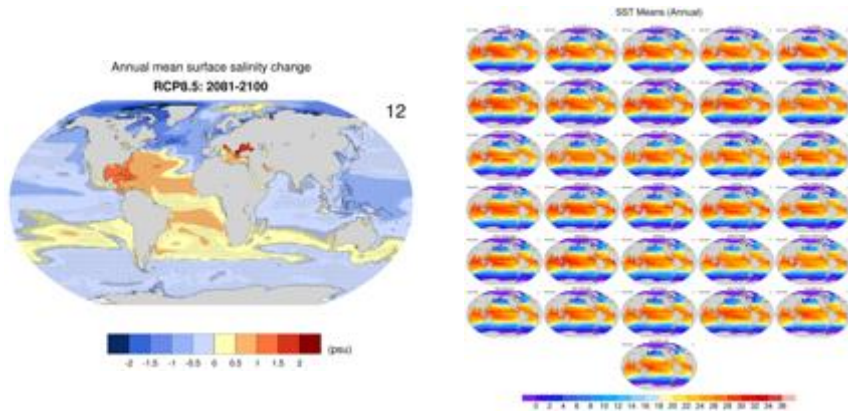


The screenshot shows the NCI Catalogue search interface. The search bar contains the query "ACCESS1-0 model output prepared for CMIP5". The results are displayed in a grid format, showing several entries for CMIP5 data. The interface includes a search bar, a "Sign in" button, and a language dropdown menu. The search results are sorted by relevancy, and the first few results are for "ACCESS1-0 model output prepared for CMIP5 1pctCO2" and "ACCESS1-3 model output prepared for CMIP5 1pctCO2". The interface also includes a sidebar with filters for "TYPE OF RESOURCES", "KEYWORDS", "PROVIDED BY", "FORMATS", and "STATUS".

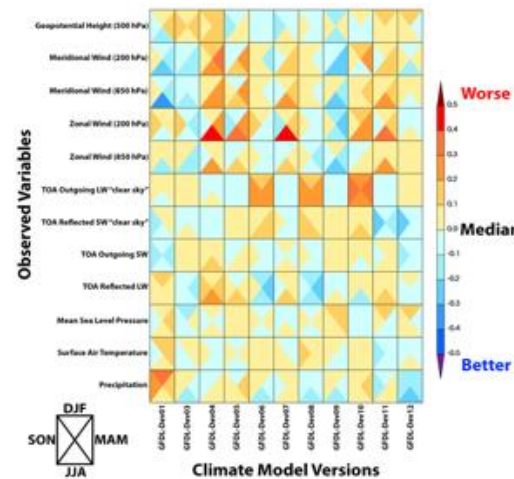
New analysis tools and updates to the existing tools

- a. Review of the CSIRO CMIP pipeline*
- b. Testing of international model evaluation tools for the CMIP6 data analysis
 - o Earth System Model eValuation Tool (ESMValTool) – diagnostics and performance metrics
 - o PCMDI Metrics Packages (PMP) – model evaluation and statistical error measures

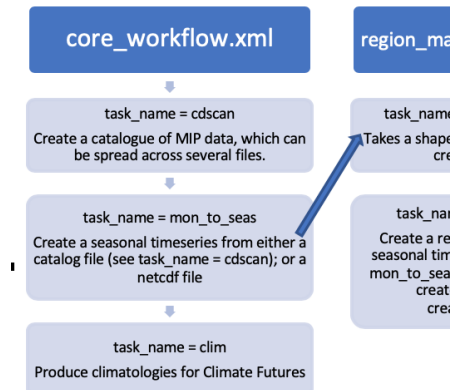
ESMValTool



PMP



CMIP Pipeline



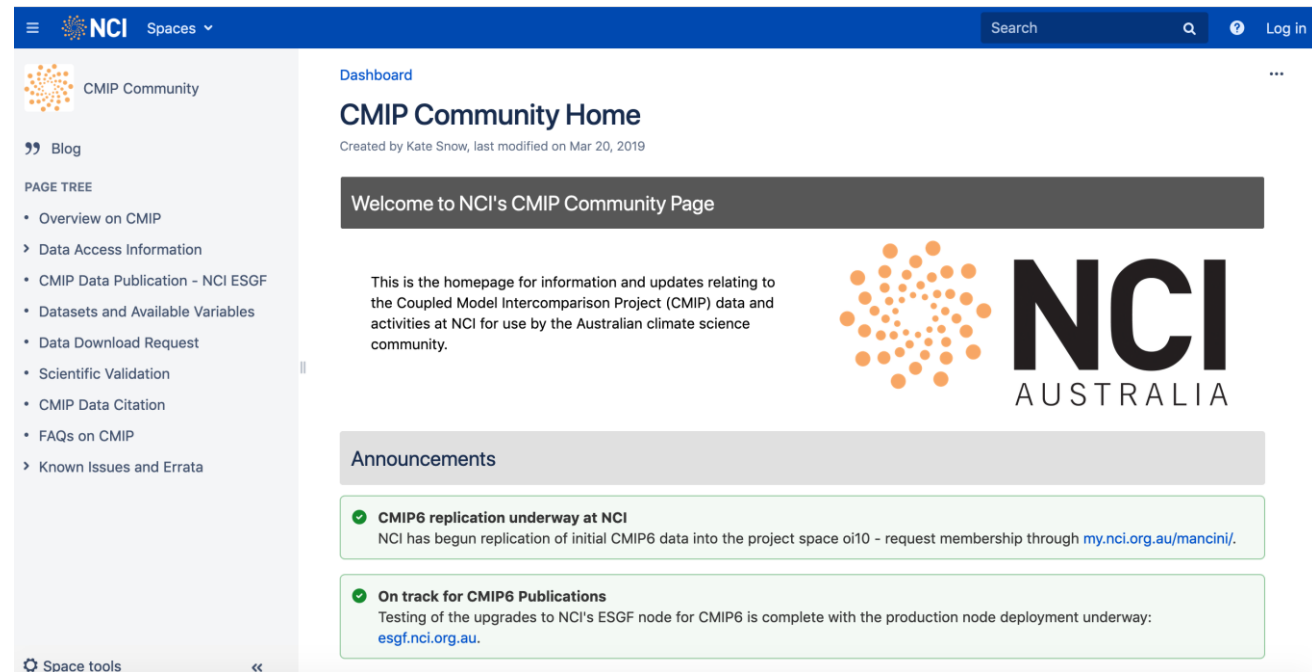
*See Claire Trenham's talk ... on now!

- Prototype of JupyterHub
 - Notebook based workflow
 - Host collection of core notebooks for climate analysis
 - Simpler training platform
 - Central administration of server environment
 - Investigation into Pangeo
- Suite of R tools and workflows
 - Most climate support focused on Python so looking to understand R user requirements.



User support - Documentation and Outreach/Training

- a. User documentation and updated CMIP Community website at NCI.
- b. New training material for the CMIP6 data and tools – to be available online.
- c. Contact help@nci.org.au to discuss training opportunities and user support.

The screenshot shows the NCI CMIP Community Home page. The header includes the NCI logo, a search bar, and a 'Log in' link. The main content area features a 'Welcome to NCI's CMIP Community Page' banner, followed by a paragraph explaining the page's purpose. Below this is an 'Announcements' section with two items: 'CMIP6 replication underway at NCI' and 'On track for CMIP6 Publications'. A left sidebar contains a 'PAGE TREE' with links to various CMIP-related resources. The footer of the page shows 'Space tools'.

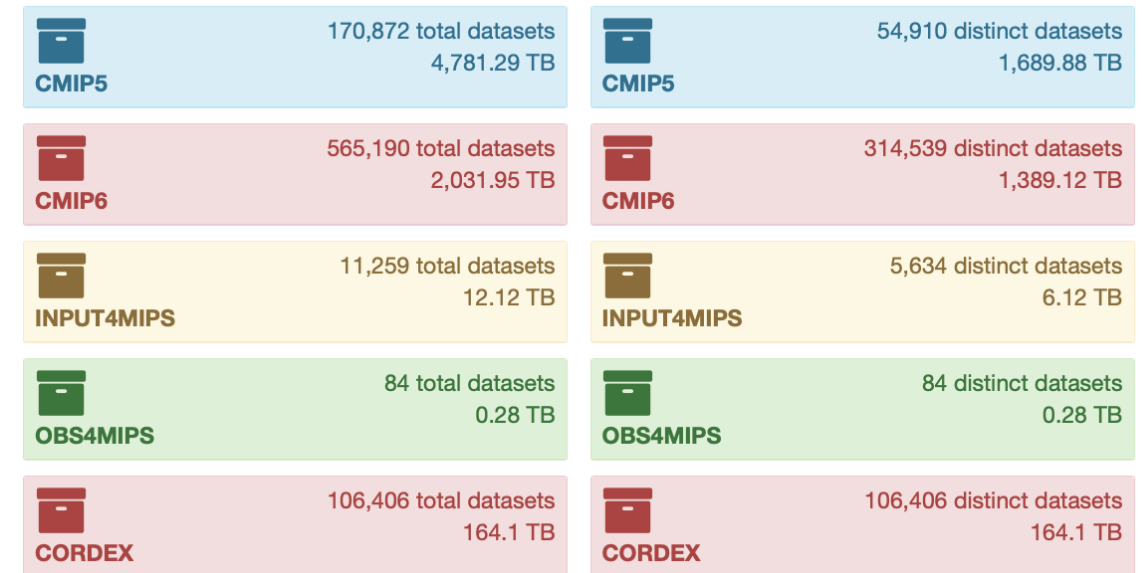
<https://opus.nci.org.au/display/CMIP/CMIP+Community+Home>

- Improved data usage and tracking
 - Improved tracking (CleF records all data searches, requests, project membership, ...) and reporting,
 - ESGF dashboard API and usage
- Development of case studies and user feedback.
 - Helps define priorities and future work requirements.
 - Demonstrates impacts.
- Establishment of ongoing sustainability requirements beyond June 2019.
 - Expected needs for CMIP6 in coming years
 - Maintaining infrastructure and services
 - Support and training

All projects



Top projects



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

Making science efficient and reliable so researchers can focus on their work

- Supporting the Australian Climate Research Community for CMIP6:
 - Data Management, replication and publishing
 - Software and Tools for data-intensive analysis
 - Improved data access
 - User support and training
 - Integration with HPC – model development and data analysis
- Some challenges along the way:
 - Download of massive datasets
 - Storage availability
- Project ends mid-2019 but new requirements are emerging:
 - New data (ERA5 &?) and research objectives
 - Increasing requirements for tools and resources
 - Establishing sustainable ongoing support



