

# Data Access Information

CMIP data located at the National Computational Infrastructure (NCI) covers a broad range of datasets, including CMIP5 and CMIP6 era replicated and published data as well as key observational and reanalysis datasets. For further information on the different climate datasets and available variables see the [Datasets and Available Variables](#) page. On this page we describe the methods in which the CMIP data hosted at NCI may be accessed. Sections in this page are broken into the following topics:

- [Access Options](#)
  - [Access on the NCI filesystem.](#)
  - [The ESGF data portal.](#)
  - [The Geonetwork catalog.](#)
- [Direct Data Access \(NCI Users Only\)](#)
- [Request to Access a Data Collection](#)
- [CleF](#)
- [Data Organisation](#)
  - [CMIP6 Published Data](#)
  - [CMIP6 Official Replica Data](#)
  - [CMIP5 and CORDEX Australian Published Data](#)
  - [CMIP5 Official Replica Data](#)
  - [CMIP3 Official Replica Data](#)
- [Guidance for Data Users](#)

Image Source: <https://www.earthsystemcog.org/projects/cmip/>

## Access Options



**Please note:** **External users** who do not have an NCI user account are directed to option 2 below. Alternatively you can find out more about [how to access NCI](#).

Access to the CMIP and related data available is available via three key methods:

- 1. Access on the NCI filesystem.**
  - a. Users with an NCI login can directly access the data locally on the NCI filesystem with both HPC [Gadi](#) and the Virtual Desktop Infrastructure ([VDI](#)).
  - b. Users are encouraged to use CleF - a python module developed for ease of searching and accessing CMIP data at NCI.
- 2. The ESGF data portal.**
  - a. [https://esgf.nci.org.au/projects/esgf\\_nci/](https://esgf.nci.org.au/projects/esgf_nci/)
  - b. The Earth System Grid Federation ([ESGF](#)) data portal provides access to CMIP datasets published by NCI and hosted across the international data nodes. If you cannot use the data on raijin or the VDI, you may search for the CMIP data from the ESGF site. Note, not all replica data is currently published on NCI ESGF site.
- 3. The Geonetwork catalog.**
  - a. You may search for the CMIP data published at NCI via the [Geonetwork](#) metadata catalog, which also provides links to data services on the THREDDS Data Service.
  - b. Australian published CMIP5 data is available on geonetwork under: [https://geonetwork.nci.org.au/geonetwork/srv/eng/catalog.search#/metadata/f3525\\_9322\\_8600\\_7716](https://geonetwork.nci.org.au/geonetwork/srv/eng/catalog.search#/metadata/f3525_9322_8600_7716).

## Direct Data Access (NCI Users Only)

Primary access to the CMIP data, from both [HPC](#) and the [VDI](#), is achieved via requesting to join the relevant project space (see below) and accessing the data from `lg/data/<project code>` (see table below).

	Australian Published CMIP Data	Replicated CMIP Data	Replicated Observational and Reanalysis Data
<b>NCI Project Code</b>	<b>fs38</b> = CMIP6-era <b>rr3</b> = CMIP5-era (incl. CORDEX)	<b>oi10</b> = CMIP6-era <b>al33</b> = CMIP5-era (incl. CORDEX) <b>cb20</b> = CMIP3	<b>qv56<sup>1</sup></b> = input4MIPs, obs4MIPs, ana4MIPs



**Please note:** New version updates for qv56 datasets will not be actioned for the remainder of 2020. Please contact [help@nci.org.au](mailto:help@nci.org.au) (Subject: Data Collections) if you have any questions about these datasets.

# Request to Access a Data Collection

You may request to join a data collection through [my.nci.org.au/mancini](http://my.nci.org.au/mancini). Your request will be sent to the data collection manager for approval and you must agree to the same Terms and Conditions that govern CMIP access data access as stipulated by the Earth Systems Grid Federation:

[CMIP6 Terms of Use](#)

[CMIP5 Terms of Use](#)

## CleF

The volume and complexity of CMIP makes manual searching for data on the filesystem a time consuming process. To efficiently find the CMIP data that you wish to access we recommend using the Climate Finder tool CleF developed by ARCCSS and CLEX. CleF is a python module designed for ease of searching and accessing data at NCI through a command line interface. Documentation on using CleF is available [here](#) and membership of any of the NCI CMIP projects oi10, al33, rr3 or ua6 is required for use (if you are a new member it may take up to 1 day for access to the NCI database used by CleF to be granted).

## Data Organisation

To permit ease of use and interdisciplinary research, CMIP uses a standard naming convention for files, directories, metadata and URLs. These conventions are described in detail in the [CMIP6 Controlled Vocabularies](#) and [CMIP5 Controlled Vocabulary](#)

### CMIP6 Published Data

Under the CMIP6 DRS, data may be found with the following directory format:

```
/g/data/fs38/publications/CMIP6/CMIP/<institution_id>/<source_id>/<experiment_id>/<member_id>/<table_id>/<variable>/<grid_label>/<version>
```

### CMIP6 Official Replica Data

Under the CMIP6 DRS, data may be found with the following directory format:

```
/g/data/oi10/replicas/CMIP6/CMIP/<institution_id>/<source_id>/<experiment_id>/<member_id>/<table_id>/<variable>/<grid_label>/<version>
```

### CMIP5 and CORDEX Australian Published Data

The CMIP5 and CORDEX published data can be found with the following directory structure:

```
/g/data/rr3/publications/CMIP5/output1/<institute>/<model>/<experiment>/<frequency>/<realm>/<table>/<ensemble>/<version>/<variable>
```

### CMIP5 Official Replica Data

The CMIP5 official replica data can be found with the following directory structure:

```
/g/data/al33/replicas/CMIP5/combined/<institute>/<model>/<experiment>/<frequency>/<realm>/<table>/<ensemble>/<version>/<variable>
```

### CMIP3 Official Replica Data

The CMIP3 official replica data can be found with the following directory structure:

```
/g/data/cb20/replicas/cmip3/<institute>/<model>/<experiment>/<frequency>/<realm>/<ensemble>/<variable>
```

Note some of the key difference in the facets between CMIP5 and CMIP6, in particular "model" in CMIP5 has become "source\_id" and "ensemble" has become "member\_id". Other facets have similar names, though in CMIP6 the convention often includes an "\_id".

Definitions for directory format terms:

- **institute/institution\_id**: the institution that produced the model output (e.g. CSIRO-BOM, UNSW, etc.)
- **model/source\_id**: also refereed as **source\_id** in CMIP6, it is the CMIP model name identifier.
- **experiment/experiment\_id**: the CMIP experiment identifier (e.g., historical, piControl, rcp45, etc.)
- **frequency**: frequency identifier (e.g., 3hr, day, mon, etc.)
- **realm**: modelling realm (e.g., atmos, land, ocean, etc.)
- **ensemble/member\_id**: ensemble or member\_id, provides information on initialisation and physics identifier (e.g., r1i1p1, r1i1p2, etc.)
- **variable/variable\_id**: output variable (see [full list here](#))
- **version**: available versions, and 'latest' with symbolic links to the latest available version, where a version isn't available the creation date is used instead.
- **table/table\_id**: For example "Amon" is short for atmosphere monthly, "Omon" is ocean monthly, etc.

## Guidance for Data Users

The following guidance material for CMIP6 data users is provided through PCMDI. It includes information on experiment design, model output, terms of use and citation, model documentation, error reporting, registering publications which use CMIP6 data, and CMIP6 governance.

<https://pcmdi.llnl.gov/CMIP6/Guide/dataUsers.html>

For more information on the CMIP and CORDEX data, including data formats and processing see [ESGF User Support - Data FAQs](#).