



# How to Search for Data

Kate Snow & Jingbo Wang

**Geonetwork**

CMIP Website

ESGF Site

CleF

Data Download Requests

- 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
- How to request to join data projects (for closed data collections)

- **A new NCI Geonetwork catalogue has been released this month!**
- The new version permits improved search options, clearer user information (such as project paths) and more detailed metadata.



The screenshot shows the NCI Geonetwork website interface. At the top, there is a search bar and a language selector set to English. A cookie consent banner is visible. Below the banner, the "Get Started" section features a search bar with the text "Search over all NCI's data collections, datasets, and services...". To the right, there is a button labeled "NCI Catalogue". The "Browse by topics" section includes buttons for "Climate/soil, national...", "Environment", "Invasive species", and "Elevation". The "Browse resources" section includes buttons for "Catalogue", "Collections", and "Default Service". The footer contains the text "Powered by 3.8.2.2" and links for "About", "GitHub", "API", and "Share on social sites".

When searching and browsing data, published collections will have a direct link to NCI's Data Services.

Datasets are stored within data subcollections and collections as parent-child relationships

Information on filesystem location immediately available



**Coupled Model Intercomparison Project (CMIP5)**

This data collection contains replicated output of the Coupled Model Intercomparison Project phase 5 (CMIP5) experiments. The latest version of CMIP5 focuses on CMIP5.

CMIP5 provides a multi-model comparison with the main aim being to improve our understanding of past, present and future climate change. Additional goals of CMIP5 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 similarly forced models produce a range of responses.

The data collection originates from the World Climate Research Programme (WCRP) (<https://www.wcrp-climate.org/ncip-cmip>) Working Group of Coupled Modelling (WGCM) with numerous climate modeling groups from around the world participating. The experiments of CMIP5 include the core B2CA experiments (including emip, 1pcrCO2, a2Control, and abrupt4CO2) as well as historical runs and other endorsed MIPs.

The international 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 able to provide a replica repository of the priority datasets. The replica data is located locally on the NCI filesystems under /gpubs/cmip5/.

For more information refer to the NCI CMIP website: <https://ncip.nci.org.au/index.php/CMIP/CMIP-home>

**Download and links**

Download at NCI: <http://ncip.nci.org.au/ncip/ncip.html>

**Associated resources**

CSIRO-BOM ACCESS1.0 model output prepared for CMIP5 (Child record)  
CMIP5 is meant to provide a framework for coordinated climate



**About this resource**

Project Code	cmip5
Local File Path	/gpubs/cmip5/publications/CMIP5/
DOI	<a href="https://doi.org/10.4225/41/5a700000000000">https://doi.org/10.4225/41/5a700000000000</a>
Keywords	<ul style="list-style-type: none"> <li>Other Chemical Sciences Environmental Chemistry</li> <li>Atmospheric Sciences</li> <li>Oceanography</li> <li>Physical Geography and Environmental Sciences</li> <li>National Computational Infrastructure NCI</li> </ul>
Classification	Unclassified
Legal constraints	The CMIP5 terms of use are found here: <a href="https://cmip5.nci.org.au/cmip5/terms.html">https://cmip5.nci.org.au/cmip5/terms.html</a>

Browse data catalogue for  
CMIP5 datasets:

<https://geonetwork.nci.org.au>

Try to searching “ACCESS1-0 model...”

Have a quick look at the type  
of information contained, the  
parent/child records, etc.



The screenshot shows the NCI Catalogue search interface. At the top, there is a search bar with the text "NCI Catalogue" and a search icon. To the right of the search bar, there are links for "Sign in" and "English". Below the search bar, a blue banner contains a cookie consent message: "This webpage uses cookies. If you continue navigating this page, we will assume you accept this. Want to know more about this message? Accept or Get me out of here".

The main content area is titled "Get Started" and includes the text "Search over all NCI's data collections, datasets, and services...". Below this is a search input field containing the text "ACCESS1-0 model". A search icon is located to the right of the input field. A dropdown menu displays the search results, with the top result highlighted in blue: "ACCESS1-0 model output prepared for CMIP5 (pControl)". Other results include "ACCESS1-0 model output prepared for CMIP5 (r10i1p0)", "ACCESS1-0 model output prepared for CMIP5 (r10i1p1)", "ACCESS1-0 model output prepared for CMIP5 (r10i1p2)", "ACCESS1-0 model output prepared for CMIP5 (r10i1p3)", "ACCESS1-0 model output prepared for CMIP5 (r10i1p4)", "ACCESS1-0 model output prepared for CMIP5 (r10i1p5)", and "CMIP5-BCR ACCESS1-0 model output prepared for CMIP5".

To the right of the search results, there is a section titled "NCI Catalogue" and a "Browse resources" section with three green buttons: "Dataset", "Collection", and "Data Service".

Geonetwork

**CMIP Website**

ESGF Site

CleF

Data Download Requests

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

Community information page on CMIP data. Includes announcements, user guides, data access information, etc.

Publicly viewable.

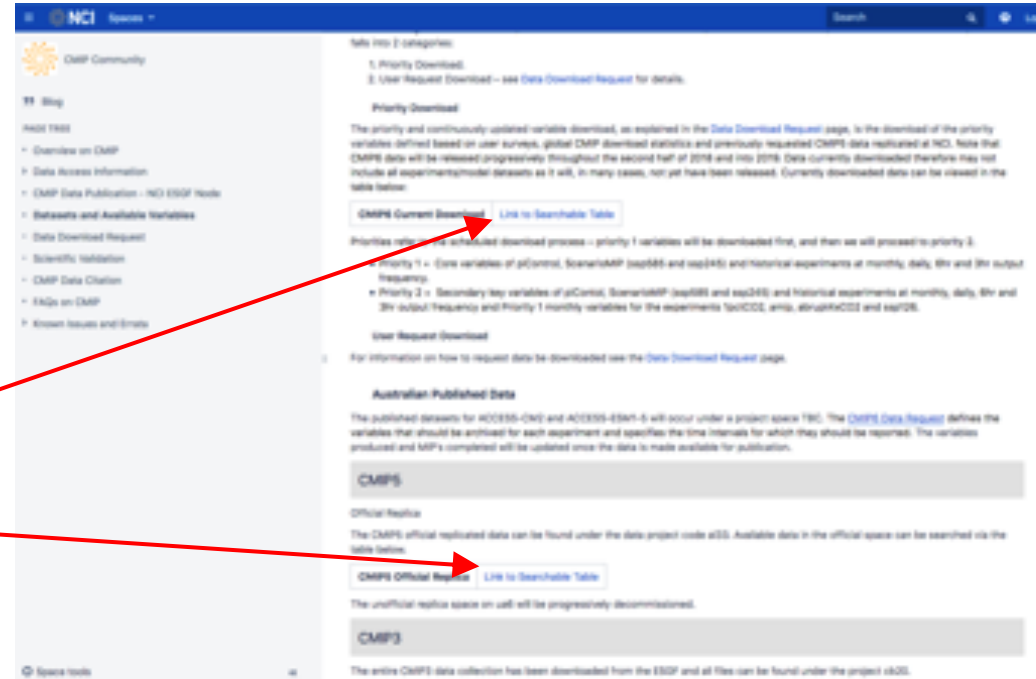
Central source of information for all CMIP relate activities at NCI.





Information on replica datasets downloaded to NCI is contained on the “Datasets and Available Variables” page.

Includes searchable tables updated weekly on data both downloaded and in the download queue.



- Table is updated weekly
- Includes download status and priority
- Relies on attribute search (more on CMIP attributes later).

### ESGF CMIP6 NCI Replication Search

Last updated 2019-05-12 19:17:21

Note that a maximum of 1000 results is displayed

Use the blue boxes below to do a query search of the data available (done) and in the download queue (queued) for CMIP6.

Note the change in attribute labels from CMIP5 to CMIP6. For example, 'model' in CMIP5 is now referred as 'source\_id' in CMIP6 and 'ensemble' in CMIP5 is 'member\_id'.

Both the download status and priorities are shown. A priority of 1-2 indicates variables in the bulk download while a priority of 'U' indicates that it has been requested by a user.

If the data you require is not found, please see the <https://opus.nci.org.au/pages/viewpage.action?spaceKey=CMIP&title=Data+Download+Request> for details on making a download request.

variable	source_id	experiment	table_id	member_id	status, priority
cl	CNRM-CM6-1	1pctCO2	Amon	r1i1p1f2	done 2
clt	CNRM-CM6-1	1pctCO2	Amon	r1i1p1f2	done 2
evspsbl	CNRM-CM6-1	1pctCO2	Amon	r1i1p1f2	done 2
hfs	CNRM-CM6-1	1pctCO2	Amon	r1i1p1f2	done 2
hfs	CNRM-CM6-1	1pctCO2	Amon	r1i1p1f2	done 2
hus	CNRM-CM6-1	1pctCO2	Amon	r1i1p1f2	done 2
pr	CNRM-CM6-1	1pctCO2	Amon	r1i1p1f2	done 2
ps	CNRM-CM6-1	1pctCO2	Amon	r1i1p1f2	done 2
pel	CNRM-CM6-1	1pctCO2	Amon	r1i1p1f2	done 2
rds	CNRM-CM6-1	1pctCO2	Amon	r1i1p1f2	done 2
rut	CNRM-CM6-1	1pctCO2	Amon	r1i1p1f2	done 2
rsds	CNRM-CM6-1	1pctCO2	Amon	r1i1p1f2	done 2

Geonetwork

CMIP Website

**ESGF Site**

CleF

Data Download Requests

- CMIP data is distributed via the Earth System Grid Federation (ESGF): a collaboration of internationally distributed data nodes.
- NCI manages the ESGF Tier 1 data repository that serves the Australian climate science community.



- Data published via ESGF can be searched from the node sites.
- <https://esgf.nci.org.au/>
- Or just Google search “NCI CMIP5/6 ESGF”

Click to go to search interface



Powered by **NCI** | Powered by **ESGF**

Welcome to the NCI node. Please report issues to [help@nci.org.au](mailto:help@nci.org.au)

Welcome, Guest | Login | Create Account

# WCRP CMIP6

World Climate Research Programme

You are at the ESGF/NCI/ESGF/NCI node

[Home](#) | [Technical Support](#)

## WCRP Coupled Model Intercomparison Project (Phase 6)

[Search & Download Data](#)

Simple Text Search

The Coupled Model Intercomparison Project, which began in 1990 under the auspices of the World Climate Research Programme (WCRP), is now in its sixth phase (CMIP6). CMIP6 coordinates somewhat independent model intercomparison activities and their experiments which have adopted a common infrastructure for collecting, organizing, and distributing output from models performing common sets of experiments. The simulation data produced by models under previous phases of CMIP have been used in thousands of research papers (some of which are listed here), and the multi-model results provide some perspective on errors and uncertainty in model simulations. This information has proved invaluable in preparing high profile reports assessing our understanding of climate and climate change (e.g., the IPCC Assessment Reports).

**Datasets from CMIP6 simulations are available through the [CMIP6 Search Interface](#)**

Also see [CMIP6 Website Summary](#) for a summary site showing available experiments and models.

The following links will lead to additional information about CMIP6:

- The [guidance for experiment \(and data sharing\) should be consulted before downloading data.](#)
- The [WCRP CMIP6](#) and [CMIP6](#) [homepage](#) pages provide background information.
- [Earth System Grid \(ESG\) 2016](#) have published a CMIP6 overview paper in a [December 2016 issue \(Development Special Issue\)](#), which also includes descriptions of each of the five [Earth System Grid \(ESG\) 2016](#). An overview of the design of the infrastructure supporting preparation and distribution of model output is provided by [Trenberth et al. \(2016\)](#).
- The [Terms of Use](#) must be followed by all who access CMIP6 model output, so please read them carefully.
- Please register on the [CMIP6 publication database](#) any articles you publish that make use of CMIP6 output.

The complete archive of CMIP6 output is made available for search and download via each of the following portals:

- USA, POC/LLNL, CORDEX – <https://esgf-node.llnl.gov/experiment/>
- France, IPSL – <https://esgf-node.jrc.ecmwf.int/experiment/>
- Germany, DLR – <https://esgf-node.dlr.de/experiment/>
- UK, UKESM – <https://esgf-node1.met.rdg.ac.uk/experiment/>

**Federated ESGF Node Sites**

- ESGF/ESG/US
- ESGF/ESG/FR
- ESGF/ESG/DE
- ESGF/ESG/UK

**Browse Projects**

Year	All	No	Tag
Phase projects (1)			
ESG/ESG			
Phase projects (2)			
ESG/ESG			
Phase projects (3)			
ESG/ESG			

Quick Search:

Not logging in via the "Tracker" may be prone to instability.

CMIP6/ESG Page Home

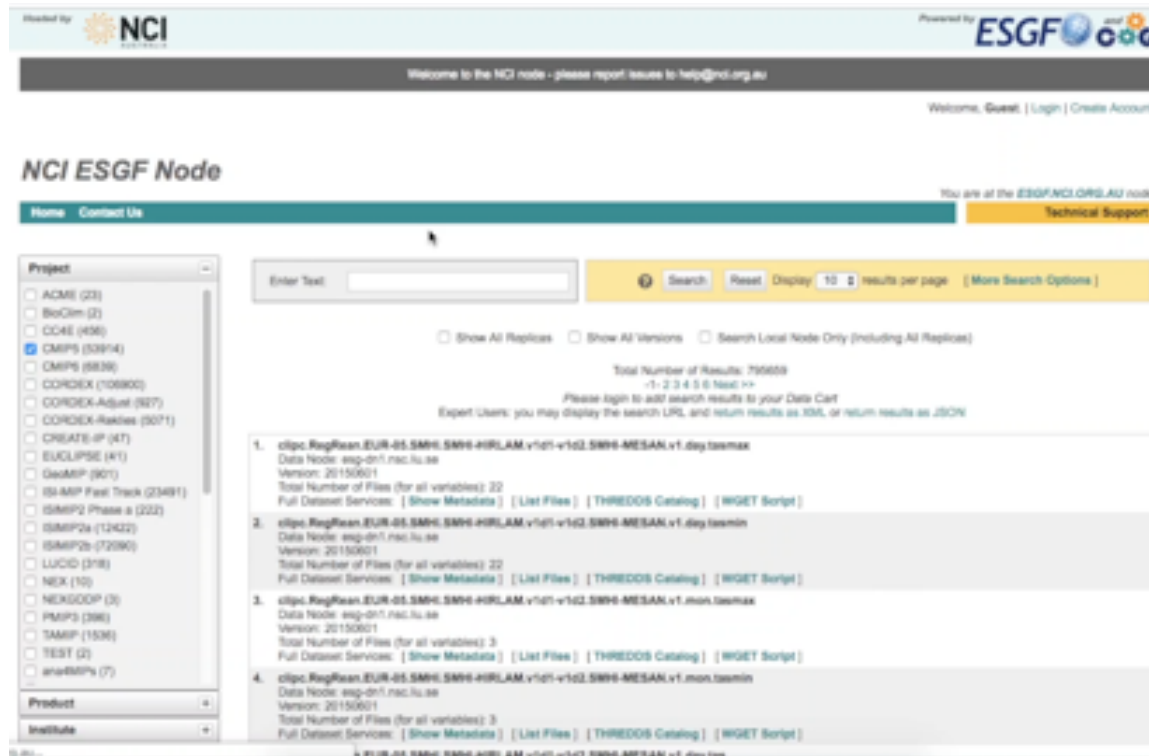
# Keyword Search

## Attribute search on LHS tabs

## Search particular node publications/replicas

## View metadata/THREDDS catalogue

## View retracted data under "Show All Versions"



The screenshot shows the NCI ESGF Node search interface. At the top, it says "Powered by NCI" and "Powered by ESGF and eosc". A welcome message reads: "Welcome to the NCI node - please report issues to help@nci.org.au". On the right, it says "Welcome, Guest | Login | Create Account".

The main heading is "NCI ESGF Node". Below it, there are navigation links for "Home" and "Contact Us", and a "Technical Support" button. A breadcrumb trail indicates "You are at the ESGF/NCI/ORG.AU node".

The search interface includes a search bar with the text "Enter text:" and a search button. To the right of the search bar, it shows "Display 10 results per page" and a link to "More Search Options". Below the search bar, there are three checkboxes: "Show All Replicas", "Show All Versions", and "Search Local Node Only (including All Replicas)".

The search results show a total of 750659 results, with a page number of 1. Below the search bar, there are instructions: "Please login to add search results to your Data Cart" and "Expert Users: you may display the search URL and return results as XML, or return results as JSON".

The search results are listed in a table with the following columns: Project, Product, and Institute. The first four results are:

- Project:** cllpc\_RegRean\_EUR-05.SM90.SM90-HIRLAM.v1r01-v1r02.SM90-MESAN.v1.day.taasmax  
**Data Node:** esg-dn1.nci.au.se  
**Version:** 20150601  
**Total Number of Files (for all variables):** 22  
**Full Dataset Services:** [ Show Metadata ] [ List Files ] [ THREDDS Catalog ] [ WGET Script ]
- Project:** cllpc\_RegRean\_EUR-05.SM90.SM90-HIRLAM.v1r01-v1r02.SM90-MESAN.v1.day.taasmin  
**Data Node:** esg-dn1.nci.au.se  
**Version:** 20150601  
**Total Number of Files (for all variables):** 22  
**Full Dataset Services:** [ Show Metadata ] [ List Files ] [ THREDDS Catalog ] [ WGET Script ]
- Project:** cllpc\_RegRean\_EUR-05.SM90.SM90-HIRLAM.v1r01-v1r02.SM90-MESAN.v1.mon.taasmax  
**Data Node:** esg-dn1.nci.au.se  
**Version:** 20150601  
**Total Number of Files (for all variables):** 3  
**Full Dataset Services:** [ Show Metadata ] [ List Files ] [ THREDDS Catalog ] [ WGET Script ]
- Project:** cllpc\_RegRean\_EUR-05.SM90.SM90-HIRLAM.v1r01-v1r02.SM90-MESAN.v1.mon.taasmin  
**Data Node:** esg-dn1.nci.au.se  
**Version:** 20150601  
**Total Number of Files (for all variables):** 3  
**Full Dataset Services:** [ Show Metadata ] [ List Files ] [ THREDDS Catalog ] [ WGET Script ]

On the left side of the interface, there is a "Project" filter menu with a list of projects and their counts. The "Show All Versions" checkbox is checked.

Geonetwork

CMIP Website

ESGF Site

**CleF**

Data Download Requests

CleF is the Climate Finder tool developed by CLEX

<https://clef.readthedocs.io/en/latest/gettingstarted.html>

Be aware that CleF is a tool under active development and changes/updates may occur.

Example to follow. Load the module.

## Getting Started

CleF is presently installed in an anaconda environment, which must be loaded before use (on either VDI or Rajat):

```
$ module use /j/ralat3/RAJAT/jupyter/modules
$ module load conda
$ source activate clef-test
```

clef is accessed through the command-line clef program. There are presently two main commands:

- `clef search` to execute searches on the CMIP5 dataset
- `clef local` to execute searches on the CMIP5 dataset

and four optional flags are available for each command to limit the output or submit a data request:

- `clef --remote search` returns all the ESGF CMIP5/CMIP6 datasets matching the constraints
- `clef --local search` Finds local files on NCI's ESGF CMIP5/CMIP6 mirror
- `clef --missing search` Finds files on ESGF that haven't been downloaded to NCI
- `clef --request search` create and pass to NCI a request to download the missing files (NB, this is not yet implemented)

If these flags are omitted then the tool will search on the ESGF datasets matching the constraints and return both the local and missing files lists.

The search works like the ESGF search website, e.g. [https://esgf.nci.org.au/search/esgf\\_nci](https://esgf.nci.org.au/search/esgf_nci). Results can be filtered by using flags matching the ESGF search facets:

```
$ clef search --model ACCESS_E1 \
  --experiment historical \
  --frequency Ann 1 \
  --variable pr \
  --variable no
```



Returning help messages to list usage and available commands/parameters.

```
File Edit View Search Terminal Help
(clef-0e1) [kur000@h1 ~]$ 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
  --debug   Show debug info
  --help    Show this message and exit.

Commands:
  cspg  Search ESGF and local database for CRPS files Constraints can be...
  cspgk Search ESGF and local database for CRPS files Constraints can be....
  ds    Search local database for non-ESGF datasets
```

```
File Edit View Search Terminal Help
(clef-0e1) [kur000@h1 ~]$ clef cspgk --help
Usage: clef cspgk [OPTIONS] [MODELS]...

Search ESGF and local database for CRPS files

Constraints can be specified multiple times, in which case they are
combined - using OR: -v tas -v tasmin will return anything matching
variable = 'tas' or variable = 'tasmin'. The --latest flag will check ESGF
for the latest version available, this is the default behaviour

Options:
  -s, --activity [AER|AOR|AORC|BGC|CHER|ISR|LAND|OCC|RAD|SLAB]
  -t, --table x          CRPS (OR) table: Amon, S05ay, S05ay ...
  -k, --model, --source_id x  CRPS model ID: GFDL-APM, CMIP-CM1 ...
  -v, --variable x       CRPS variable name as in filenames
  -id, --header TEXT      CRPS header ID: <sub-exp-id>-RELDATE
  -g, --grid, --grid_label TEXT  CRPS grid label: i.e. gn for the model
  --active-grid
  -r, --resolution, --nominal_resolution TEXT
  Approximate resolution: '250 km', pass in
  quotes
```



Geonetwork

CMIP Website

ESGF Site

CleF

**Data Download Requests**

Once you have completed a search using the CleF tool, a data download request may be made if you have determined the data to be of high priority and required for your research.

To submit a request for the data to be downloaded, add “--request”

So from previous example, it becomes:

```
$ clef--request cmip6 --activity CMIP -e historical -v tasmin -v ccb -mi r1i1p1f1
```

Requests from the VDI will be emailed to [help@nci.org.au](mailto:help@nci.org.au). Requests from Raijin need to be manually emailed.

You can see the data being downloaded using CleF which also contains the information on data in the download queue.

```
File Edit View Search Terminal Help
(clef-test) [kxs900@vdi-n20 ~]$ clef cmip5 -e rcp85 -v uss -m EC-EARTH
/g/data1b/a133/replicas\DRIPS\combined/ICHEC/EC-EARTH/rcp85/mon/atmos/Amon/r711p1/v20171115/uss/
/g/data1b/a133/replicas\DRIPS\combined/ICHEC/EC-EARTH/rcp85/mon/atmos/Amon/r811p1/v20171115/uss/

The following datasets are not yet available in the database, but they have been requested or recently do
wnloaded
cmip5_output1_ICHEC_EC-EARTH_rcp85_mon_atmos_Amon_r1411p1_v20171115_uss status: queued
```

Further CleF training: <https://climateextremes.org.au/cms-training-clef-discovering-cmip6/>

You can also request to have data downloaded from the CMIP Community Site

Ensure you have searched for the data **BEFORE** making a download request

Can request CMIP5, CORDEX and CMIP6 data.



How to Request CMIP Data at NCI

Steps to request NCI download CMIP or related data for Australian climate science research:

- 1. Check if the desired dataset is already available**
  - a. First check if the dataset is on the file system using [DataF](#) or by checking the table of currently downloaded data via the searchable table on the [Datasets and Available Variables](#) page.

PCMDI have released a webpage tracking [ESGF CMIP6 Data Holding](#), which shows number of datasets currently available globally for model and experiment.
- 2. If the dataset is not on the file system, check if it is in the download plan**
  - a. The user checks details on dataset download plan

[CMIP6 Priority Download](#) [Link to Searchable Table](#)

  - b. If the user sees their data in the download plan with the data to be provided in a suitable time frame, no further action is required.
- 3. If required data is not on the file system or in the download plan**
  - a. If a user wishes to request data to be downloaded at NCI for analysis, please fill out the relevant request form below. In each case you will need to include:
    - i. Variables required and the output table, if with frequency and realm,
    - ii. Models to be included,
    - iii. The experiment and ensemble information,
    - iv. Your institution and NCI user name,
    - v. The requested data will be added to the download queue with details of priority and download status.

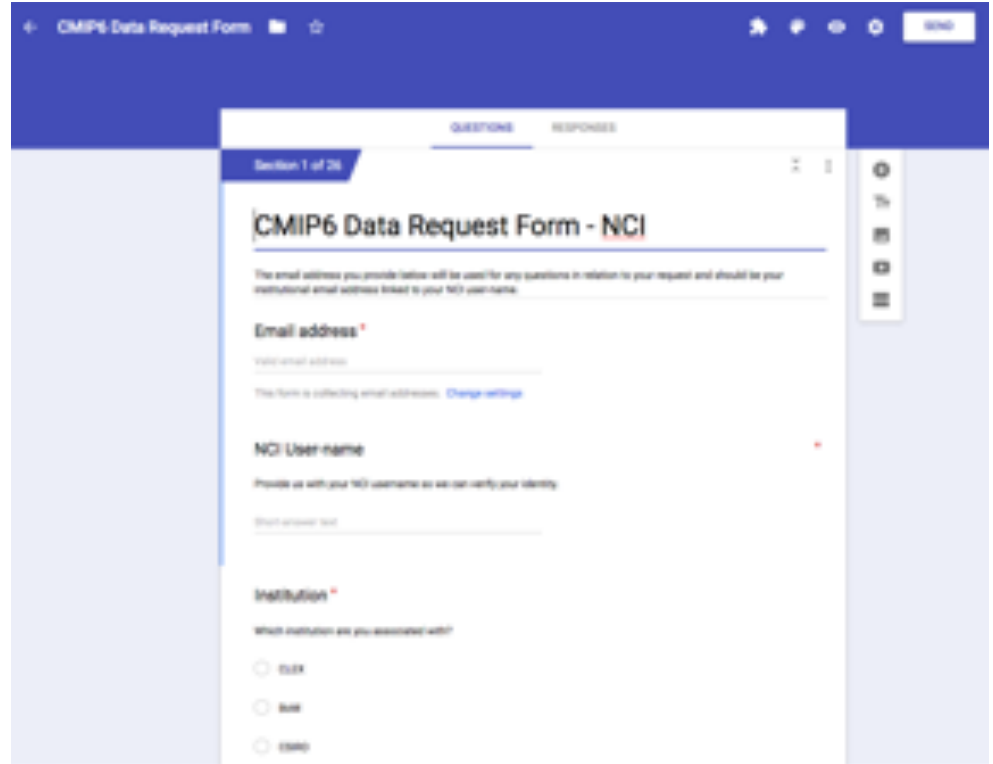
To request that CMIP5, CMIP6 or CORDEX data be downloaded at NCI, please use the following respective forms:

  - [CMIP5 Replica Download Request Form](#)
  - [CMIP6 Replica Download Request Form](#)
  - [CORDEX-CMIP6 Replica Download Request Form](#)

(or alternatively you may submit a request via the [DataF tool](#).)

Request is via a Google Forms sheet that is emailed to [help@nci.org.au](mailto:help@nci.org.au) once complete.

Time for download may vary depending on how many other downloads are in the queue, how large the download is, and where it is coming from (some nodes are slower than others).

A screenshot of a Google Forms interface titled "CMIP6 Data Request Form - NCI". The form is displayed within a browser window. The form content includes a title, a note about email addresses, and three required fields: "Email address", "NCI User name", and "Institution". The "Institution" field has three radio button options: "CSIRO", "ANU", and "UNSW". The browser's address bar shows "CMIP6 Data Request Form" and the page has a blue header with navigation icons and a "SEND" button.