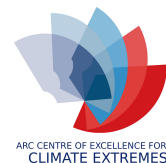
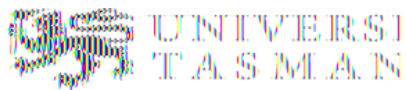




# Intro to NCI Data Collections

Data Science Week 2021

Special thanks to all supporters:



# Key Takeaways:

1. How to search and browse the NCI Data Catalogue
2. How to access the reference collections and datasets at NCI

# NCI Data Collections

- Host Australia's largest research collection of climate, weather, Earth systems, environmental, satellite, and geophysics research datasets.
- Data is a mix of nationally generated datasets as well as replicated international datasets that need to be hosted at NCI for local analysis.
- There are currently more than 13 PB of nationally and internationally significant datasets, and growing...
- As well as data being available more generally, one of the important aspects about this data is that it is organised next to high performance computing and data analysis systems.

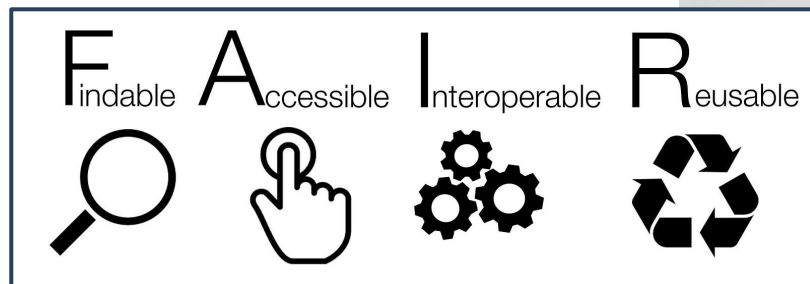


# NCI Data Collections

NCI supports a number of key internationally recognised data principles:

- **FAIR** data principles for its major data collections. FAIR is Findable, Accessible, Interoperable, Reusable.
- Programmable and high performance access
- Open as possible, Closed as necessary
- Use Data Standards wherever possible
- Transdisciplinary access

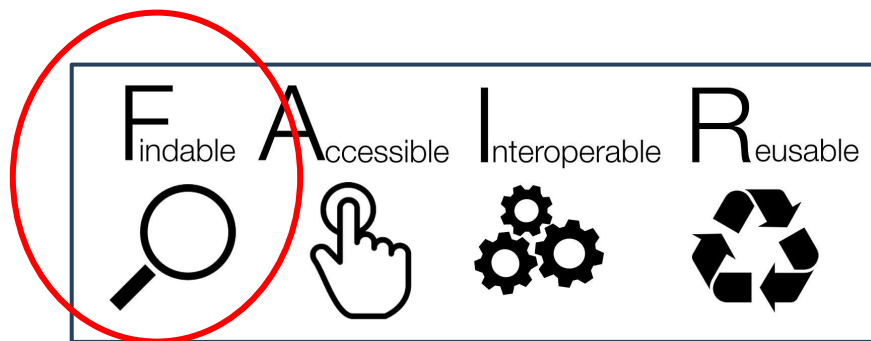
More background can be found on our wiki: <https://opus.nci.org.au/x/NAOVAg>



[Image credit: SangyaPundir](#)



# Browsing and searching for data at NCI



[Image credit: SangyaPundir](#)

# NCI Data Catalogue

While the Data Catalogue is open and many datasets are available through public services, a number of data collections are only accessible on the NCI systems (i.e., [Gadi](#), [VDI](#)) or are more useful within the NCI computational systems. This requires you to complete the NCI [new user registration](#) prior to requesting access to a data collection project.

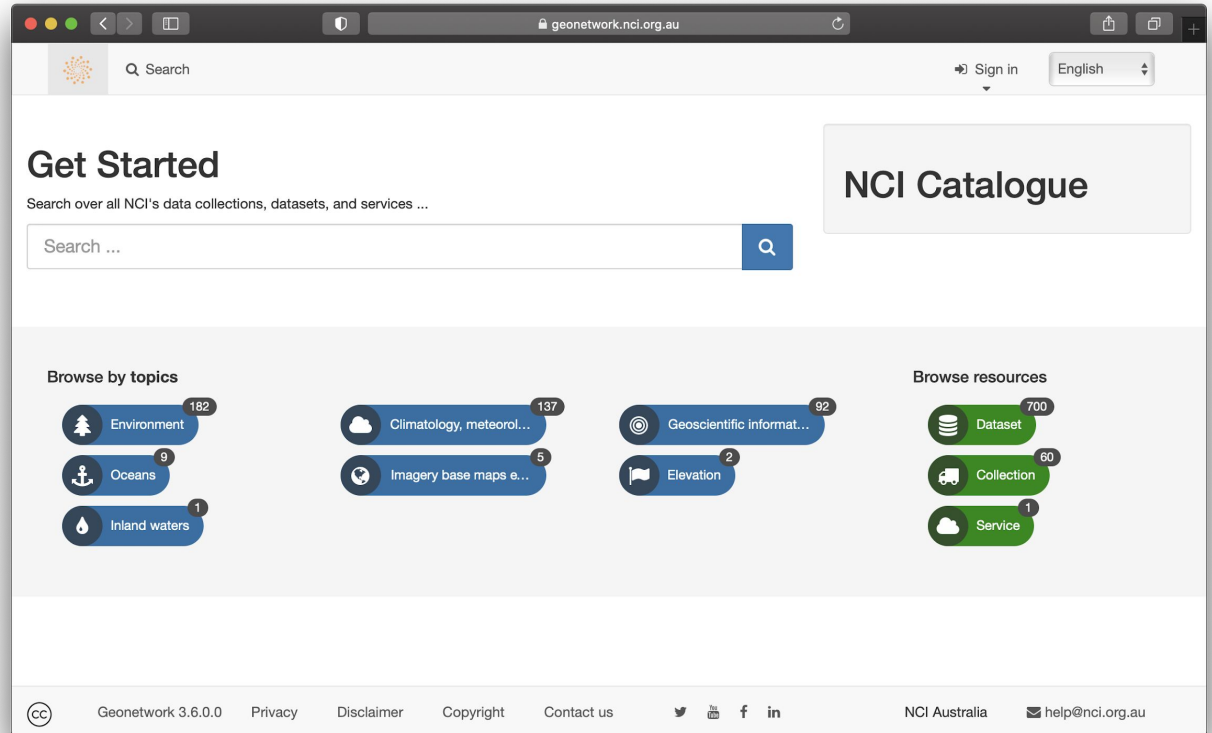
[Data Catalogue User Guide](#)



# NCI Data Catalogue

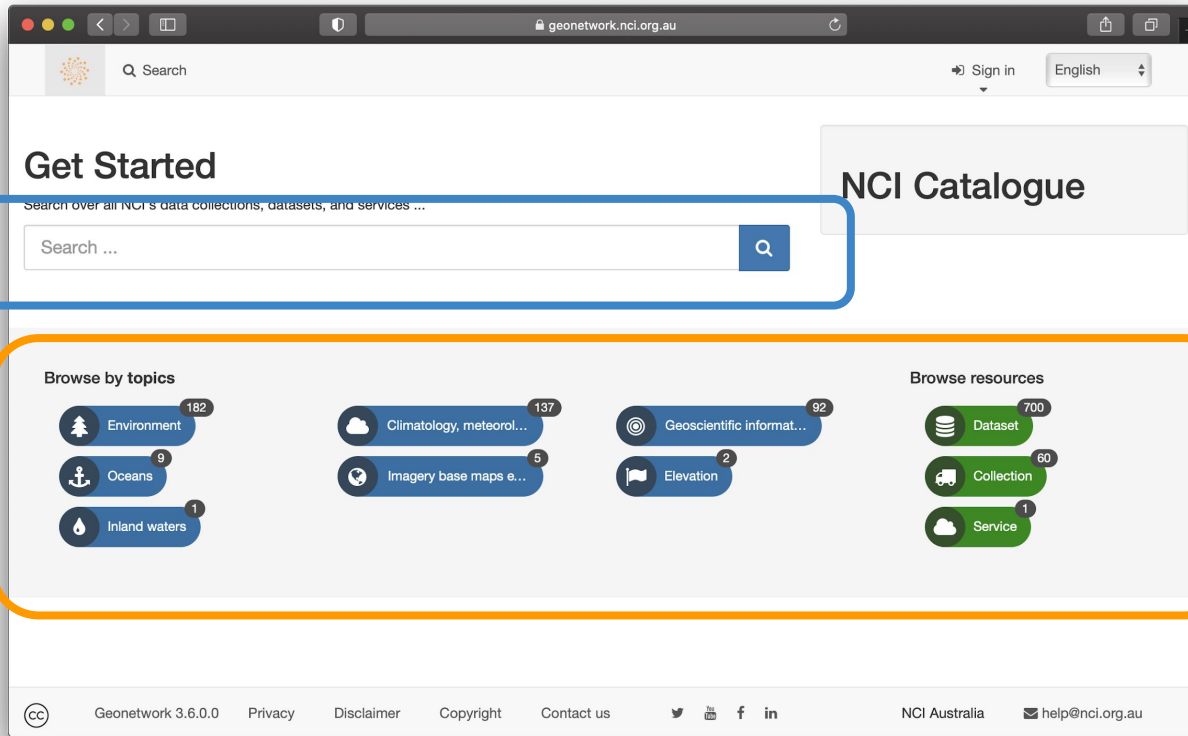
You can find information about the NCI data holdings in our Data Catalogue:

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



# Browsing and searching datasets

Search



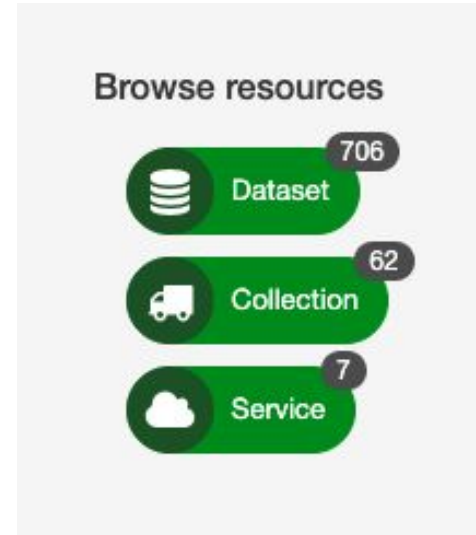
Browse

# Browsing and searching datasets

The screenshot shows the Geonetwork NCI Australia website interface. At the top, there is a search bar with the text "Search ..." and a magnifying glass icon. 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 large "Get Started" section with the text "Search over all NCI's data collections, datasets, and services ..." and another search input field. To the right of this section is a prominent "NCI Catalogue" button. Below the search bar, there are two main sections: "Browse by topics" and "Browse resources". The "Browse by topics" section features three columns of topic buttons: Environment (182), Oceans (9), and Inland waters (1) in the first column; Climatology, meteorol... (137) and Imagery base maps e... (5) in the second column; and Geoscientific informat... (92) and Elevation (2) in the third column. The "Browse resources" section, which is highlighted with an orange rounded rectangle, contains three buttons: Dataset (700), Collection (60), and Service (1). At the bottom of the page, there is a footer with a Creative Commons license icon, version information "Geonetwork 3.6.0.0", and links for "Privacy", "Disclaimer", "Copyright", and "Contact us". Social media icons for Twitter, YouTube, Facebook, and LinkedIn are also present, along with the text "NCI Australia" and the email address "help@nci.org.au".

# Browsing and searching datasets

Several types of resources you can browse by



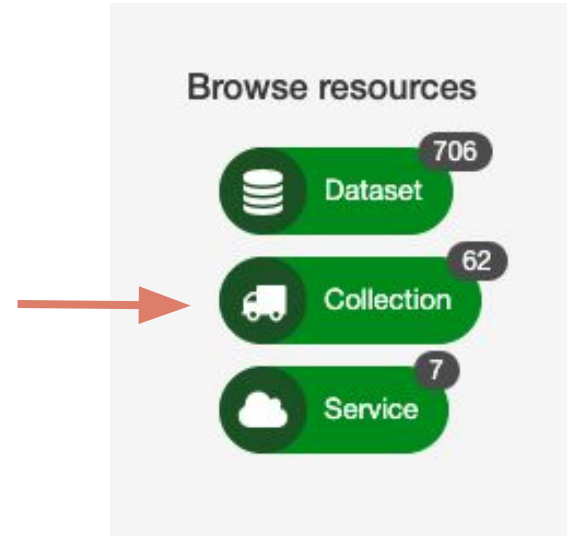
# NCI Data Collections

- A **data collection** is the highest in the hierarchy of data groupings at NCI. It is comprised of either an exclusive grouping of data subcollections; or, it is a tiered structure with an exclusive grouping of lower tiered data collections, where the lowest tier data collection will only contain data subcollections.
- A **data subcollection** is an exclusive grouping of datasets (i.e., belonging to only one subcollection) where the constituent datasets are tightly managed. It must have responsibilities within one organization with responsibility for the underlying management of its constituent datasets. A data subcollection constitutes a strong connection between the component datasets, and is organized coherently around a single scientific element (e.g., model, instrument). A subcollection must have compatible licenses such that constituent datasets do not need different access arrangements.
- A **dataset** is a compilation of data that constitutes a programmable data unit that has been collected and organized using a self-contained process. For this purpose it must have a named data owner, a single license, one set of semantics, ontologies, vocabularies, and has a single data format and internal data convention. A dataset must include its version.
- A **dataset granule** is used for some scientific domains that require a finer level of granularity (e.g., in satellite Earth Observation datasets). A granule refers to the smallest aggregation of data that can be independently described, inventoried, and retrieved as defined by NASA. Dataset granules have their own metadata and support values associated with the additional attributes defined by parent datasets.

# Collections

These records will:

- Tell you the information relevant to all datasets or products within a collection.
- They will also list these products within the record under a section 'Associated Records'



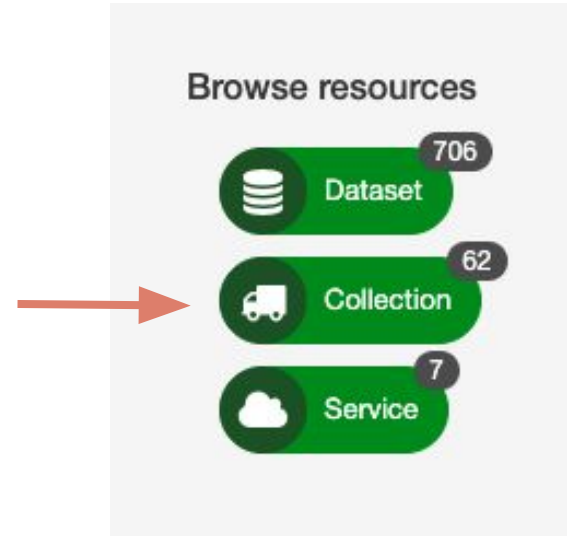
# Collections

## Good for:

- High-level browsing, especially for collections with large number of datasets

## Drawback:

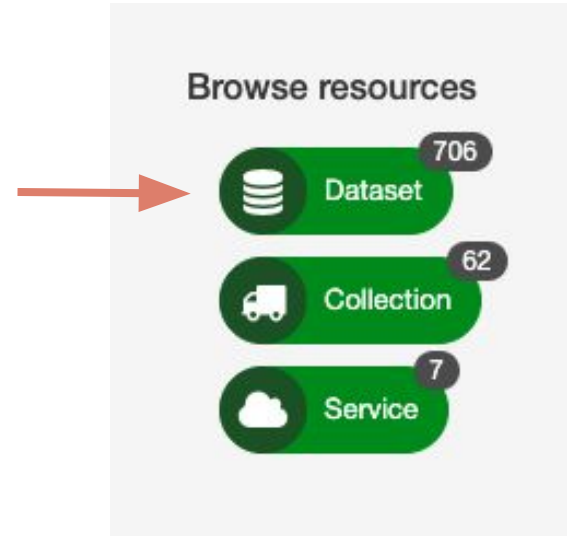
- Dataset specific information will not be displayed on these pages



# Datasets

Data records will contain more detailed information about the data product (or sometimes groups of products). This will include metadata items such as:

- Abstract
- Data access
- Licensing
- Data lineage
- Data format





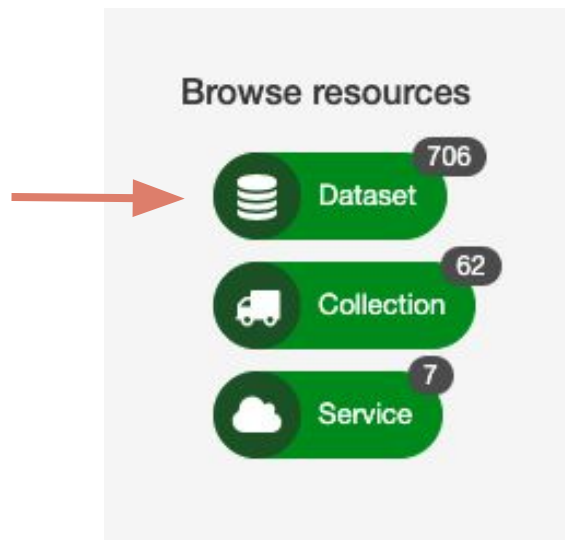
# Datasets

## Good for:

- Dataset specific information such as data access, dois, versioning, data lineage, etc.

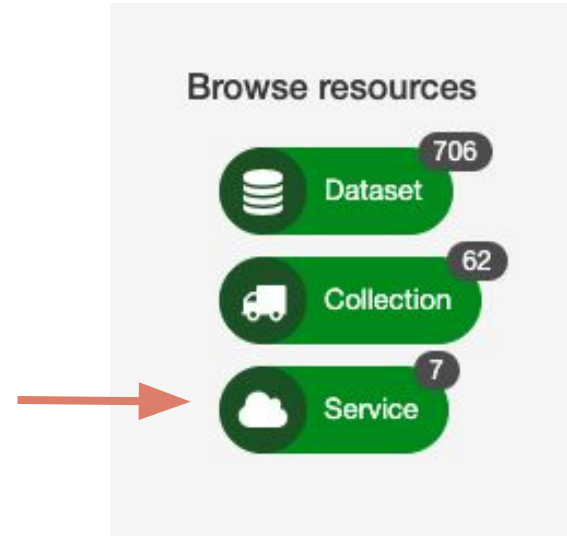
## Drawback:

- Lots of them so harder to browse.
- Geonetwork search function has some limitations.



# Services

- Service records contain information about a data service provided by NCI.
- These records will link to background information about the service and can be used when citing data access, systems or environments at NCI.



# Example 1: Browsing the ERA5 Data Collection

The screenshot shows a web browser window at [geonetwork.nci.org.au](http://geonetwork.nci.org.au). The page features a search bar with the text "era5" entered. A dropdown menu displays a list of search results under the heading "ERA5 Replicated Datasets". To the right, there is a "Browse resources" section with three categories: "Dataset" (700 items), "Collection" (60 items), and "Service" (1 item). The page also includes a "Get Started" section and an "NCI Catalogue" button.

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

era5

**ERA5 Replicated Datasets**

- ERA5 hourly data on pressure levels
- ERA5 hourly data on single levels
- ERA5 monthly averaged by hour of day data on pressure levels
- ERA5 monthly averaged by hour of day data on single levels
- ERA5 monthly averaged data on pressure levels
- ERA5 monthly averaged data on single levels
- ERA5-Land Replicated Datasets
- ERA5-Land monthly averaged data
- ERA5-Land monthly averaged data by hour of day

92

**Browse resources**

- Dataset 700
- Collection 60
- Service 1

Sign in English

CC Geonetwork 3.6.0.0 Privacy Disclaimer Copyright Contact us NCI Australia help@nci.org.au

# Record metadata

The screenshot shows a web browser window displaying the record metadata for 'ERA5 Replicated Datasets' on the Geonetwork NCI Australia website. The page includes a search bar, navigation links, and a 'Download' button. The main content area is divided into sections for the record title, abstract, and data access. The record title is 'ERA5 Replicated Datasets', and the abstract describes the dataset as the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF). The data access section provides links for local access and the NCI local file path. The footer contains the Geonetwork 3.6.0.0 logo, privacy and disclaimer links, and contact information for NCI Australia.

geonetwork.nci.org.au

Search

Sign in English

Back to search < Previous Next >

Download

**ERA5 Replicated Datasets** ← Record title

View Metadata XML

**Creation date**  
2020-08-28

**Publication date**  
2020-11-16

**Revision date**  
2020-11-16

ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF). Data at NCI is replicated through the Copernicus Climate Data Store (CDS) and hosted for use by the Australian research community.

This collection consists of monthly and sub-daily products from 1979 (soon to be backdated to 1950) to present time (approx. 3 months in arrears), including the ERA5.1 re-run for years 2000-2006. Access and location details of these products can be found in the linked catalogue entries under the Associated Records section below or on the NCI ERA5 Community site.

Full details of the model and outputs can be found in the ECMWF ERA5 Data Documentation.

← Record abstract

**Data Access**

**Register for local access** <https://my.nci.org.au/mancini/project/rt52/join>

**NCI local file path** /g/data/rt52/era5

NCI ERA5 Community Site  
<https://opus.nci.org.au/x/R4DYB> Open link

ECMWF ERA5 Data Documentation  
<https://confluence.ecmwf.int/display/CKB/ERA5%3A+A+data+documentation> Open link

Geonetwork 3.6.0.0 Privacy Disclaimer Copyright Contact us

NCI Australia  
help@nci.org.au

# Record metadata

**ERA5 Replicated Datasets**

ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF), through the Copernicus Climate Data Store (CDS) and the Australian research community.

This collection consists of monthly and sub-daily products backdated to 1950 to present time (approx. 3 months). ERA5.1 re-run for years 2000-2006. Access and location can be found in the linked catalogue entries under the Associated Records or on the NCI ERA5 Community site.

Full details of the model and outputs can be found in the Documentation.

**Data Access**

Register for local access <https://my.nci.org.au>

NCI local file path /g/data/rt52/era5

NCI ERA5 Community Site <a href="https://opus.nci.org.au/x/R4DYB">https://opus.nci.org.au/x/R4DYB</a>	Open link
ECMWF ERA5 Data Documentation <a href="https://confluence.ecmwf.int/display/CKB/ERA5%3A+A+data+documentation">https://confluence.ecmwf.int/display/CKB/ERA5%3A+A+data+documentation</a>	Open link

**Associated Records**

ERA5.1 hourly data on pressure levels (Child record) ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF). ERA5.1 is a re-run of ERA5, for the years 2000 to 2006 only, and was... more...	Child record
ERA5.1 monthly averaged by hour of day data on pressure levels (Child record) ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF). ERA5.1 is a re-run of ERA5, for the years 2000 to 2006 only, and was... more...	Child record
ERA5.1 monthly averaged data on pressure levels (Child record) ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF). ERA5.1 is a re-run of ERA5, for the years 2000 to 2006 only, and was... more...	Child record

**Annotations:**

- Data registration link: <https://my.nci.org.au/mancini/project/rt52/join>
- Data location at NCI: /g/data/rt52/era5
- Related sites & documentation: <https://opus.nci.org.au/x/R4DYB> and <https://confluence.ecmwf.int/display/CKB/ERA5%3A+A+data+documentation>
- Link to remote data access (when available): <https://confluence.ecmwf.int/display/CKB/ERA5%3A+A+data+documentation>
- Linked parent/child records such as the sub-collection(s) and dataset(s) within a collection: ERA5.1 hourly data on pressure levels (Child record), ERA5.1 monthly averaged by hour of day data on pressure levels (Child record), ERA5.1 monthly averaged data on pressure levels (Child record)

Geonetwork 3.6.0.0 | Privacy | Disclaimer | Copyright | Contact us | NCI Australia | help@nci.org.au

# Record metadata

Search

Q Back to search < Previous Next >

## ERA5 Replicated Datasets

ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF) through the Copernicus Climate Data Store (CDS) and the Australian research community.

This collection consists of monthly and sub-daily products backdated to 1950 to present time (approx. 3 months). ERA5.1 re-run for years 2000-2006. Access and location can be found in the linked catalogue entries under the Associated Records section on the NCI ERA5 Community site.

Full details of the model and outputs can be found in the ERA5 Data Documentation.

**Data Access**

**Register for local access** <https://my.nci.org.au/mar>

**NCI local file path** /g/data/rt52/era5

**NCI ERA5 Community Site** <https://opus.nci.org.au/x/R4DYB>

**ECMWF ERA5 Data Documentation** <https://confluence.ecmwf.int/display/CKB/ERA5+A+data+documentation>

**Data Access**

**Register for local access** <https://my.nci.org.au/mar>

**NCI local file path** /g/data/rt52/era5

**NCI ERA5 Community Site** <https://opus.nci.org.au/x/R4DYB>

**ECMWF ERA5 Data Documentation** <https://confluence.ecmwf.int/display/CKB/ERA5+A+data+documentation>

**Associated Records**

- ERA5.1 hourly data on pressure levels (Child record)**  
ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF). ERA5.1 is a re-run of ERA5, for the years 2000 to 2006 only, and was... [more...](#)
- ERA5.1 monthly averaged by hour of day data on pressure levels (Child record)**  
ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF). ERA5.1 is a re-run of ERA5, for the years 2000 to 2006 only, and was... [more...](#)
- ERA5.1 monthly averaged data on pressure levels (Child record)**  
ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF). ERA5.1 is a re-run of ERA5, for the years 2000 to 2006 only, and was... [more...](#)

**Record Overview**

**NCI project code** rt52

**Catalogue record DOI** 10.25914/5f48874388857  
URL: <http://dx.doi.org/10.25914/5f48874388857>

**Categories** Climatology, meteorology, atmosphere

**Field of Research** Atmospheric Sciences [ 0401 ]

**Access constraints** License to use Copernicus products (<https://cds.climate.copernicus.eu/cdsapp/#/terms/licence-to-use-copernicus-products>)  
All users of Copernicus Products must provide clear and visible attribution to the Copernicus programme. Please see section 5 of the license for further detail.

**Dataset Information**

**Update frequency** Continual

**Format** NetCDF4

**Lineage** The replicated ERA5 datasets at NCI are downloaded from the Climate Data Store (<https://cds.climate.copernicus.eu/cdsapp#/home>). The replication steps are annotated in the 'history' attribute within each netcdf file metadata.

**Credit** The availability of ERA5 data at NCI has been made possible through ARC LIEF Grant LE200100040. The LIEF grant is supported by a collaboration and co-funding involving Monash University, the Australian National University, the University of Melbourne, the University of New South Wales, the University of Tasmania and NCI.

Geonetwork 3.6.0.0 Privacy Disclaimer Copyright Contact us NCI Australia help@nci.org.au

Persistent URL to this record

Data licensing and terms of use

# Record metadata

Search results for ERA5 Replicated Datasets. The page shows a search bar, navigation links, and a list of datasets. A blue box highlights the description of ERA5: "ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF) through the Copernicus Climate Data Store (CDS) and the Australian research community." Below this, it states: "This collection consists of monthly and sub-daily products backdated to 1950 to present time (approx. 3 months). ERA5.1 re-run for years 2000-2006. Access and location can be found in the linked catalogue entries under the Associated Records section." A blue box also highlights the "Data Access" section, which includes a "Register for local access" link and the "NCI local file path" `/g/data/rt52/era5`.

Two overlapping screenshots showing the "Data Access" and "Associated Records" sections. The "Data Access" section includes a "Register for local access" link, the "NCI local file path" `/g/data/rt52/era5`, and links to the "NCI ERA5 Community Site" and "ECMWF ERA5 Data Documentation". The "Associated Records" section lists three records: "ERA5.1 hourly data on pressure levels (Child record)", "ERA5.1 monthly averaged by hour of day data on pressure levels (Child record)", and "ERA5.1 monthly averaged data on pressure levels (Child record)". Each record includes a brief description and a "more..." link.

Record Overview for NCI project code `rt52`. The page displays various metadata fields: "Catalogue record DOI" (10.25914/544874388857), "Categories" (Climatology, meteorology, atmosphere), "Field of Research" (Atmospheric Sciences [ 0401 ]), and "Access constraints" (License to use Copernicus product). The "Dataset Information" section shows "Update frequency" as "Continual" and "Format" as "NetCDF4". The "Lineage" section describes the replicated ERA5 datasets. The "Credit" section provides information about the availability of ERA5 data at NCI, mentioning the ARC LIEF Grant LE200100040 and the support from Monash University, the Australian National University, the University of Melbourne, the University of New South Wales, and the University of Tasmania and NCI.

Contact Information section for the record. It includes a "Download metadata" link, the "Owner" (European Centre for Medium-Range Weather Forecasts (ECMWF) and Australian Research Council), and the "Publisher" (NCI Australia). The "Help" section provides contact information for questions about the record. An orange arrow points to the "Owner" information, with the text "Information on the resource owner(s), author(s), funder(s)".

# Example: Browsing ERA5

NetCDF4 (10)

**UPDATE FREQUENCIES**

Continual (12)

## ERA5-Land Replicated Datasets

ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF). ERA5-Land is a land surface dataset forced by ERA5 atmospheric parameters but with no additional data assimilation. Data at NCI is replicated through the Copernicus Climate Data Store (CDS) and hosted for use by the Australian research community. This collection consists monthly and sub-daily products from 1981 (soon to be backdated to 1950) to present time (approx. 3 months in arrears). Access and location details can be found in the linked catalogue entries under the Associated Records section below or on the NCI ERA5 Community site. Full details of the model and outputs can be found in the ECMWF ERA5 Data Documentation.

## ERA5-Land monthly averaged data

ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF). ERA5-Land is a land surface dataset forced by ERA5 atmospheric parameters but with no additional data assimilation. This dataset includes monthly averaged data for ERA5-Land and has been replicated through the Copernicus Climate Data Store (<https://doi.org/10.24381/cds.68d2bb30>). Full details of the model and outputs can be found in the ECMWF ERA5 Data Documentation.

## ERA5-Land monthly averaged data by hour of day

ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF). ERA5-Land is a land surface dataset forced by ERA5 atmospheric parameters but with no additional data assimilation. This dataset includes monthly averaged by hour of day data for ERA5-Land and has been replicated through the Copernicus Climate Data Store (<https://doi.org/10.24381/cds.68d2bb30>). Full details of the model and outputs can be found in the ECMWF ERA5 Data Documentation.

## ERA5 Replicated Datasets

ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF). Data at NCI is replicated through the Copernicus Climate Data Store (CDS) and hosted for use by the Australian research community. This collection consists of monthly and sub-daily products from 1979 (soon to be backdated to 1950) to present time (approx. 3 months in arrears), including the ERA5.1 re-run for years 2000-2006. Access and location details of these products can be found in the linked catalogue entries under the Associated Records section below or on the NCI ERA5 Community site. Full details of the model and outputs can be found in the ECMWF ERA5 Data Documentation.





# ERA5 Community Site

Further information can be found here, including:

- Data availability (down to specific variables)
- Data organisation
- Replication plan
- Known local issues

**ERA5 monthly averaged data on pressure levels** [View Metadata XML](#)

**Creation date**  
2020-11-15

**Publication date**  
2020-11-16

**Revision date**  
2020-11-16

ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF).

This dataset includes monthly averaged data on pressure levels and has been replicated through the Copernicus Climate Data Store (<https://doi.org/10.24381/cds.6860a573>)

Full details of the model and outputs can be found in the ECMWF ERA5 Data Documentation.

**Data Access**

**Register for local access** <https://my.nci.org.au/mancini/project/rt52/join>

**NCI local file path** /g/data/rt52/era5/pressure-levels/monthly-averaged

**NCI ERA5 Community Site** <https://opus.nci.org.au/x/R4DYB>

**ECMWF ERA5 Data Documentation** <https://confluence.ecmwf.int/display/CKB/ERA5%3A+A+data+documentation>

Link to data provider's data documentation

Geonetwork 3.6.0.0 Privacy Disclaimer Copyright Contact us NCI Australia [help@nci.org.au](mailto:help@nci.org.au)

# ERA5 Community Site

Further information can be found here, including:

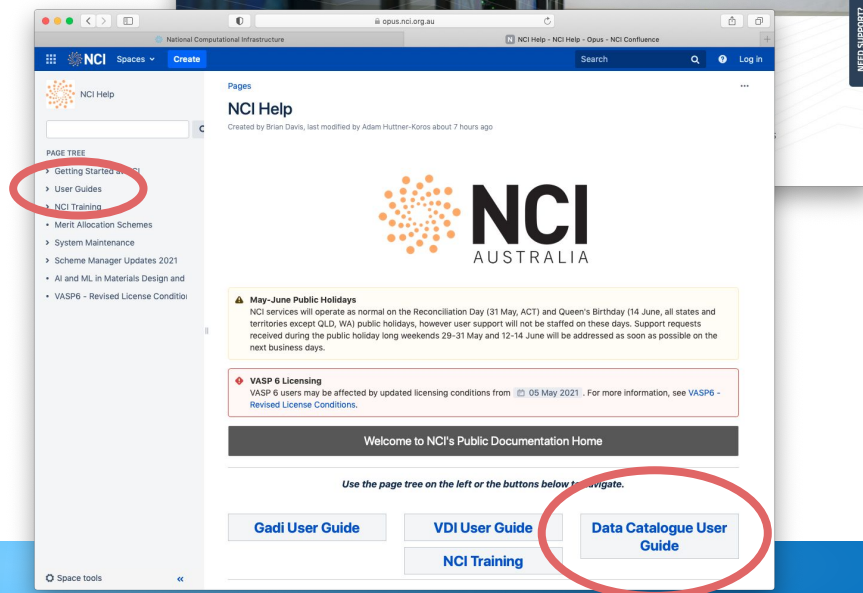
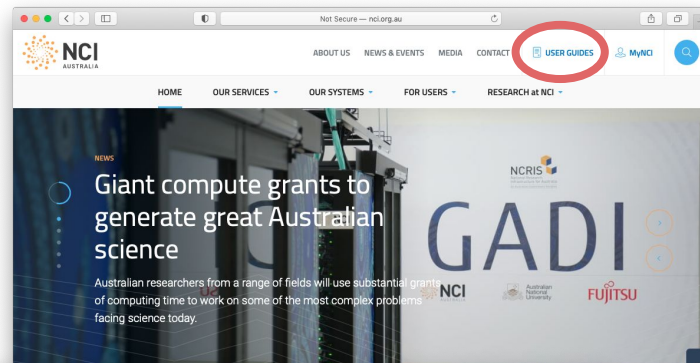
- Data availability (down to specific variables)
- Data organisation
- Replication plan
- Known local issues

The screenshot shows a web browser displaying the ERA5 Community Home page. The browser's address bar shows 'opus.nci.org.au'. The page header includes the NCI logo, 'Spaces', and a 'Create' button. A search bar and 'Log in' link are also present. The main content area features a 'Pages' section with the title 'ERA5 Community Home', created by Clare Richards and last modified by Kelsey Druken on Nov 12, 2020. A dark banner reads 'Welcome to NCI's ERA5 Community Page'. Below this, a paragraph states: 'This is the homepage for information and updates relating to the ERA5 reanalysis data and activities at NCI for use by the Australian research community.' A list of links includes 'ERA5 at NCI', 'Data Access' (with sub-links for 'Dataset Availability', 'Data Organisation', and 'Known Issues'), 'ERA5 Parameter Shortname Exceptions', and 'FAQs on ERA5'. To the right, the NCI AUSTRALIA logo is displayed. Two notification boxes are present: a green one for '16 Nov 2020 ERA5 Initial Collection Release. To register for access please see: Data Access.' and a yellow one for 'Replication Status' with information on available or upcoming datasets. The 'Acknowledgements' section follows, detailing the ARC LIEF Grant LE200100040 and the supporting institutions: Monash University, Australian National University, The University of Melbourne, UNSW, University of Tasmania, and the Australian Government/Australian Research Council. At the bottom, a user comment says 'Chris Wilkinson likes this' and a footer provides contact information: 'For any questions or assistance please contact help@nci.org.au.' The page also indicates 'No labels'.

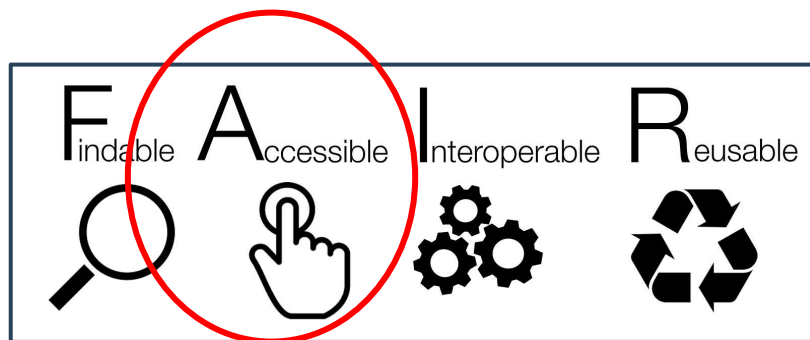
# Exercise 1

- From the NCI homepage: <https://nci.org.au/>, select the **User Guides** link in the top menu
- Select the **Data Catalogue User Guide**
- Take a moment to browse some of our collections, datasets and services.

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



# Accessing data



[Image credit: SangyaPundir](#)

# Accessing data

Several ways you can access the NCI Data Collections:

1. Directly on the NCI systems:

- Gadi (HPC)
- VDI (Virtual Desktop)

2. Remotely through data services

# Direct at NCI

- To access the Data Collections directly at NCI, you'll first need to be a registered user with system resources (this is usually through your organisation or institution).
- For more information on getting started at NCI, please go to our main webpage: <https://nci.org.au/users/how-access-nci>
- Any questions on how to get started, please reach out to us at [help@nci.org.au](mailto:help@nci.org.au)

# Direct at NCI

- Each data collection is hosted within its own project space.
- Users need to register for access to each of these spaces.
  - These registration links are in each of the data catalogue entries.
  - You will be directed to our MyNCI user portal where you will need to agree to the license and terms of use for the collection.

ERA5 monthly averaged data on pressure levels

ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF).

This dataset includes monthly averaged data on pressure levels and has been replicated through the Copernicus Climate Data Store (<https://doi.org/10.24381/cds.6860a573>)

Full details of the model and outputs can be found in the ECMWF ERA5 Data Documentation.

Creation date  
2020-11-15

Publication date  
2020-11-16

Revision date  
2020-11-16

**Data Access**

**Register for local access** <https://my.nci.org.au/mancini/project/rt52/join>

NCI local file path: /g/data/rs2/era5/pressure\_levels/monthly\_averaged

NCI ERA5 Community Site  
<https://opus.nci.org.au/x/R4DYB>

ECMWF ERA5 Data Documentation  
<https://confluence.ecmwf.int/display/CKB/ERA5%3A+A+data+documentation>

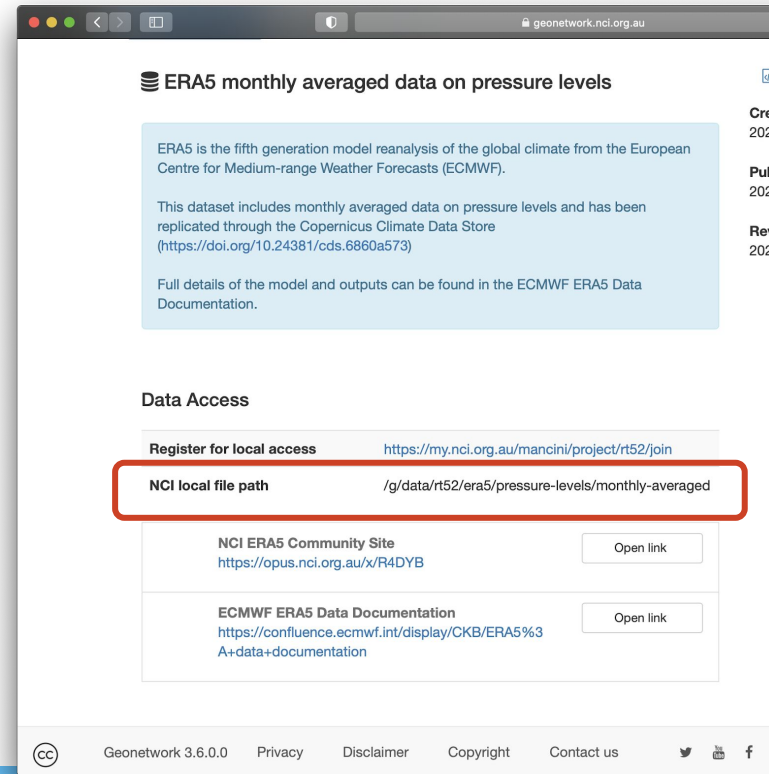
Geonetwork 3.6.0.0 Privacy Disclaimer Copyright Contact us

# Navigating to the data from Gadi or VDI

Once registered as a member of the project collection code, you will be able to view and access the collection contents from both Gadi and the VDI.

## To navigate to the data:

- Locate the collection or dataset location from the catalogue record
- Login to Gadi or the VDI as per usual (please see our [User Guides](#) here if you would like more information on getting started to either of these systems)
- Use the 'cd' command to move to the data location:
  - `cd <dataset_path>`



The screenshot shows a web browser window with the URL `geonetwork.nci.org.au`. The page title is "ERA5 monthly averaged data on pressure levels". The main content area contains a light blue box with the following text:

ERA5 is the fifth generation model reanalysis of the global climate from the European Centre for Medium-range Weather Forecasts (ECMWF).

This dataset includes monthly averaged data on pressure levels and has been replicated through the Copernicus Climate Data Store (<https://doi.org/10.24381/cds.6860a573>)

Full details of the model and outputs can be found in the ECMWF ERA5 Data Documentation.

Below this is a "Data Access" section with a table:

Register for local access	<a href="https://my.nci.org.au/mancini/project/rt52/join">https://my.nci.org.au/mancini/project/rt52/join</a>
<b>NCI local file path</b>	<code>/g/data/rt52/era5/pressure-levels/monthly-averaged</code>
NCI ERA5 Community Site <a href="https://opus.nci.org.au/x/R4DYB">https://opus.nci.org.au/x/R4DYB</a>	<input type="button" value="Open link"/>
ECMWF ERA5 Data Documentation <a href="https://confluence.ecmwf.int/display/CKB/ERA5%3A+A+data+documentation">https://confluence.ecmwf.int/display/CKB/ERA5%3A+A+data+documentation</a>	<input type="button" value="Open link"/>

The "NCI local file path" row is highlighted with a red border. At the bottom of the page, there is a footer with the Creative Commons logo, version "Geonetwork 3.6.0.0", and links for "Privacy", "Disclaimer", "Copyright", "Contact us", and social media icons for Twitter, YouTube, and Facebook.



# Accessing data from your Gadi compute job

- When accessing data from your Gadi job you will need to add the data collection code within the '**-l storage**' [PBS flag](#).
  - This requirement is new to Gadi.

```
#!/bin/bash

#PBS -l ncpus=48

#PBS -l mem=190GB

#PBS -l jobfs=200GB

#PBS -q normal

#PBS -P a00

#PBS -l walltime=02:00:00

#PBS -l storage=gdata/a00+scratch/a00

#PBS -l wd
```

# Data Services at NCI

NCI hosts and supports a wide range of data services:

1. General data servers used for a wide range of collections:
  - a. [THREDDS](#) - Quick overview here in next few slides
  - b. [GSKY](#) - Link to user guide can be found in this catalogue record
2. Community specific data services:
  - a. [NCI ESGF \(Earth System Grid Federation\) Data Portal](#)
  - b. [Copernicus SARA Data Portal](#)
  - c. [Optical Astronomy data services](#)

# 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 access methods

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 browser window. The form includes sections for Action, Data URL, Global Attributes, Variables, and Remote access. A red arrow points to the Data URL field, which contains a long URL. The Global Attributes section shows metadata for a NetCDF file. The Variables section lists 'y' as an array of 64-bit reals. The Remote access section notes that the URL can be used in place of local file paths. At the bottom, a large data array is displayed, consisting of many rows 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). Not available on all collections.

The screenshot displays the NetCDF Subset Service (NCSS) web interface. The main form is titled "NCSS for Grids ( Grid as Point Dataset )". It shows the following details:

- Dataset:** /thredds/ncss/rs0/tiles/EPSG3577/LS8\_OLI\_TIRS\_NBAR/LS8\_OLI\_TIRS\_NBAR\_3577\_-10\_-21\_2013.nc ( Dataset )
- Description:** [Red arrow points to this field]
- Base Time:** 2013-03-26T01:44:16.825Z [Green arrow points to this field]
- Select Variable(s):** [Red arrow points to this field]
- Variables with available Times:** A long list of variables and their corresponding time ranges, such as "band\_1 = Nadir BRDF Adjusted Reflectance 0.43-0.45 microns (Coastal Aerosol)".
- Choose Spatial Subset:** A map showing the selected location with a bounding box. [Green arrow points to the map]
- Choose Lat/Lon Location:** A form with fields for Latitude (-18.8) and Longitude (123). [Green arrow points to the Latitude field]
- time extents:** A section for selecting time ranges, with "Start" set to 2013-03-26T01:44:16.825Z and "End" set to 2013-12-27T01:51:20.926Z. [Green arrow points to the Start field]
- output format:** A dropdown menu set to "csv". [Green arrow points to the dropdown]
- variables:** A list of variables with checkboxes, where "band\_3" and "band\_5" are selected. [Green arrow points to the "band\_3" checkbox]
- NCSS Request URL:** A text area containing the full request URL, including parameters like "var=band\_4&var=band\_5&var=band\_7&lat=123&lon=123&time\_start=2013-03-26T01:44:16.825&time\_end=2013-12-27T01:51:20.926&accept=csv".

Green arrows on the right side of the image point to various elements: "latitude, longitude" points to the map and location fields; "time extents" points to the time selection fields; "output format" points to the dropdown menu; and "variables" points to the variable selection list.

# New releases

- All new releases are included in the monthly NCI Newsletters as well as in our [Data Catalogue User Guide](#).
- Major release announcements are sent to registered NCI users.

NCI regularly publishes new and updated data collections. This list provides a summary of recent updates to our [data catalogue](#).

## April 2021

- COSIMA ACCESS-OM2 0.1 degree global model output (interannual forcing simulation)
- ACCESS NWP APS3 Operational Reference Data Collection
- ACCESS NWP APS2 Operational Reference Data Collection (now available on NCI THREDDS Data Server)
- Derived Optimal Linear Combination Evapotranspiration - DOLCE v3.0
- EC-Earth3 SSP585 atmospheric forcing dataset for the Ice Algae Model Intercomparison Project (IAMMIP) experiment v1.0
- Bluelink Ocean Reanalysis - BRAN2020
- ESGF CMIP6 (new datasets added)
- ERA5 (new datasets added)

## March 2021

- Seasonal Prediction ACCESS-S2 Hindcast (1981-2018) and Supporting Data Assimilation and Intercomparison Project (SDAIP)
- High-Resolution Modelling of Extreme Storms over the East Coast of Australia v1.0
- Himawari-8 GeoCat 1.0.3 Australian Domain
- ERA5 Replicated Datasets (new products added)
- ESGF CMIP5 and CMIP6 (new datasets added)
- Geoscience Australia Geophysics Reference Data Collection (new datasets added)

**Q&A**