



Counting on 'COUNTER RD': Investigating Standardised Usage Metrics for Data at NCI

eResearch Australasia

Wednesday 30th October 2024

Jo Croucher, Andrew Robinson, Lesley Wyborn, Yiling Liu, Hannes Hollmann, Nigel Rees,
Ben Evans & Rebecca Farrington

We acknowledge and celebrate the First Australians on whose traditional lands we meet and pay our respect to our Elders past, present and future

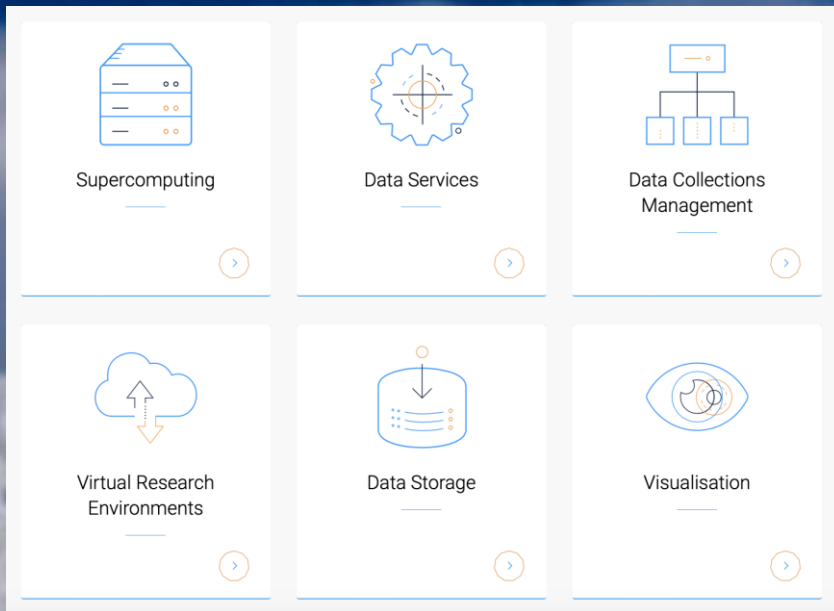


National Computational Infrastructure (NCI Australia)

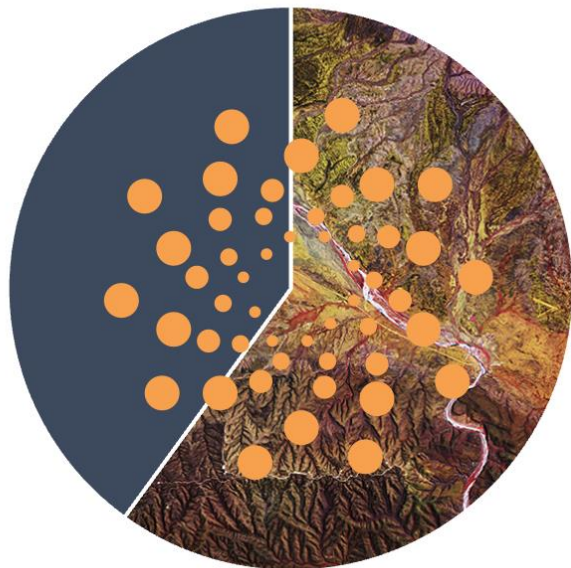


NCI Services

Range of platforms & services & data access points



Why host datasets at NCI?



- Storage for **BIG** data
- Efficiency of **co-locating data & compute**
- **Managed** datasets

*60% of NCI compute used
for climate & weather, earth &
environmental science*

Why track impact?

- Support **evidenced-based** data management decisions that support the needs of the community.
- Demonstrate **return on investment** to stakeholders.



- Making national high-resolution geophysics reference collections suitable for exascale computing in 2030.
- Collaborative project involving NCI, AuScope, TERN, and ARDC.
- The project makes rawer, high-resolution versions of AuScope-funded data accessible online and integrated with existing government datasets hosted at NCI, TERN and elsewhere.



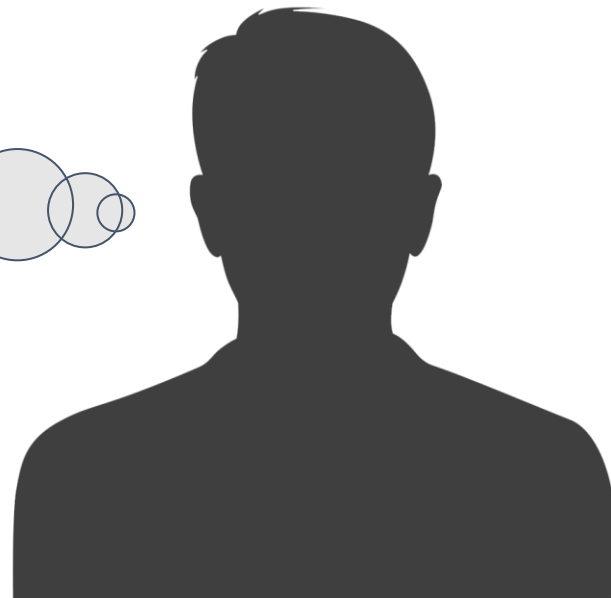


What is COUNTER?



Citations

Measuring impact





**COUNTER
METRICS**



Standardised

USAGE

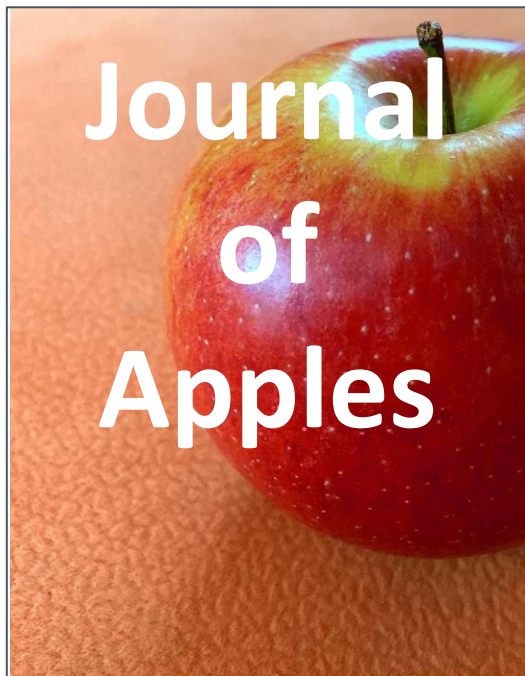
of electronic
resources



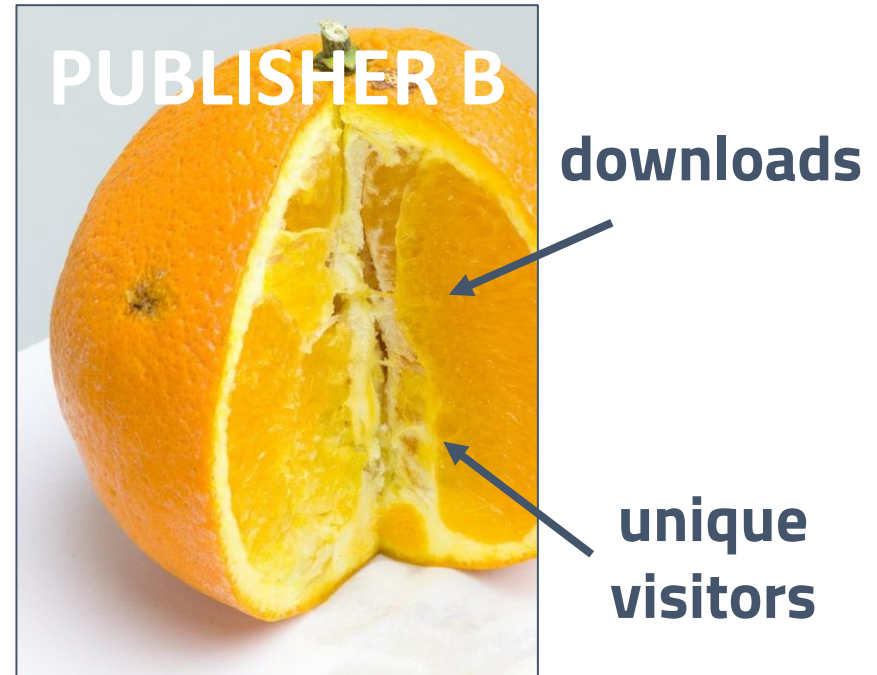
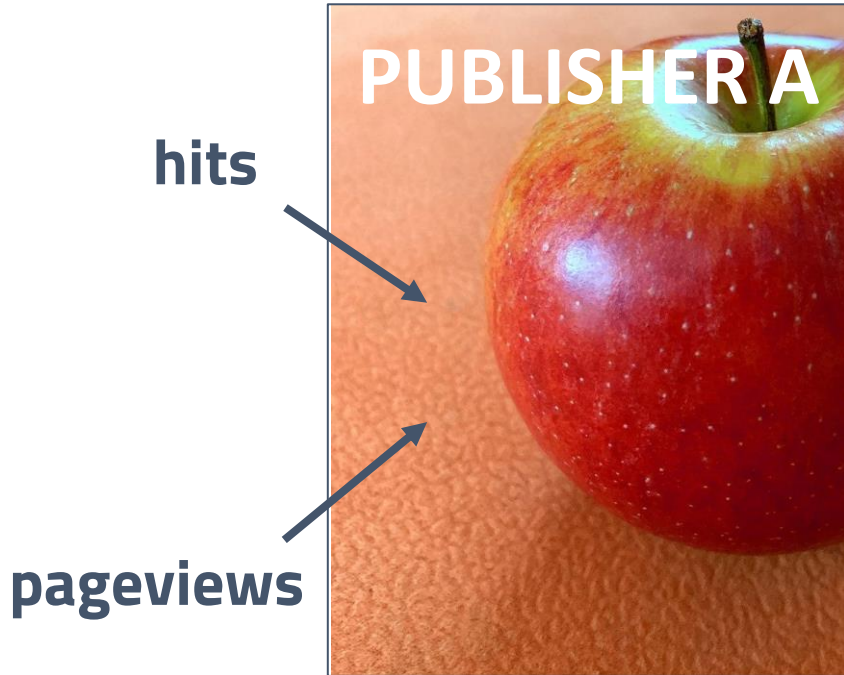
Investigations
(e.g. preview/
abstract)

Requests
(e.g. download)

Comparing usage

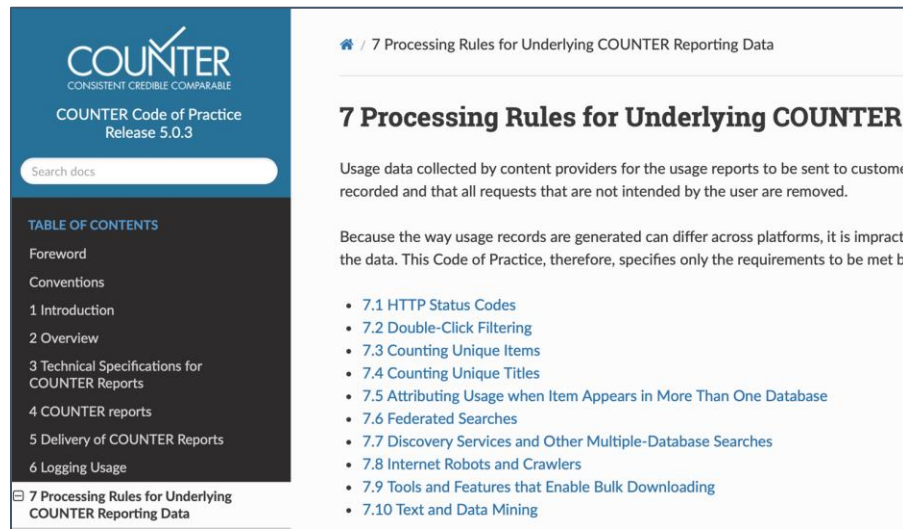


Non-standard metrics



COUNTER Code of Practice

- ✓ Standard definitions
- ✓ Detailed technical specifications (for logging usage, processing logs & sharing standard reports)
- ✓ Audited



The screenshot displays the COUNTER Code of Practice website. The header includes the COUNTER logo with the tagline 'CONSISTENT CREDIBLE COMPARABLE' and the text 'COUNTER Code of Practice Release 5.0.3'. Below the header is a search bar labeled 'Search docs'. A 'TABLE OF CONTENTS' sidebar lists various sections, with '7 Processing Rules for Underlying COUNTER Reporting Data' highlighted. The main content area shows the title '7 Processing Rules for Underlying COUNTER' and a paragraph explaining that usage data collected by content providers for reports must be sent to customers and that requests not intended by the user are removed. A list of 10 sub-sections follows, including HTTP Status Codes, Double-Click Filtering, Counting Unique Items, and Text and Data Mining.

7 Processing Rules for Underlying COUNTER Reporting Data

7 Processing Rules for Underlying COUNTER

Usage data collected by content providers for the usage reports to be sent to customer recorded and that all requests that are not intended by the user are removed.

Because the way usage records are generated can differ across platforms, it is impractical the data. This Code of Practice, therefore, specifies only the requirements to be met by

- 7.1 HTTP Status Codes
- 7.2 Double-Click Filtering
- 7.3 Counting Unique Items
- 7.4 Counting Unique Titles
- 7.5 Attributing Usage when Item Appears in More Than One Database
- 7.6 Federated Searches
- 7.7 Discovery Services and Other Multiple-Database Searches
- 7.8 Internet Robots and Crawlers
- 7.9 Tools and Features that Enable Bulk Downloading
- 7.10 Text and Data Mining

COUNTER Code of Practice



COUNTER METRICS

Delivering The Standard For Usage Metrics

Known as the Code of Practice, the COUNTER standard means that publishers, aggregators and technology providers can deliver credible, consistent, comparable usage metrics to libraries and consortia around the world.

About Us

Code Of Practice



Code of Practice



Education



FAQs



About

COUNTER Code of Practice for Research Data

- ✓ Tracks “investigations”
(e.g., data catalogue HTML views)
- ✓ Tracks number and volume of “requests”
(e.g., data downloads or data views/access)
- ✓ Documents how to manage and process log files & share standardised reports





NCI COUNTER RD TRIAL

COUNTER RD trial

1. Obtain sample [NCI THREDDS](#) web log files
2. Generate COUNTER CSV mock-up
3. Map log entries to COUNTER report elements
4. Install [COUNTER Processor application](#)
5. Configure COUNTER processor using sample THREDDS log
6. Generate sample COUNTER report for selected Geophysics 2030 datasets

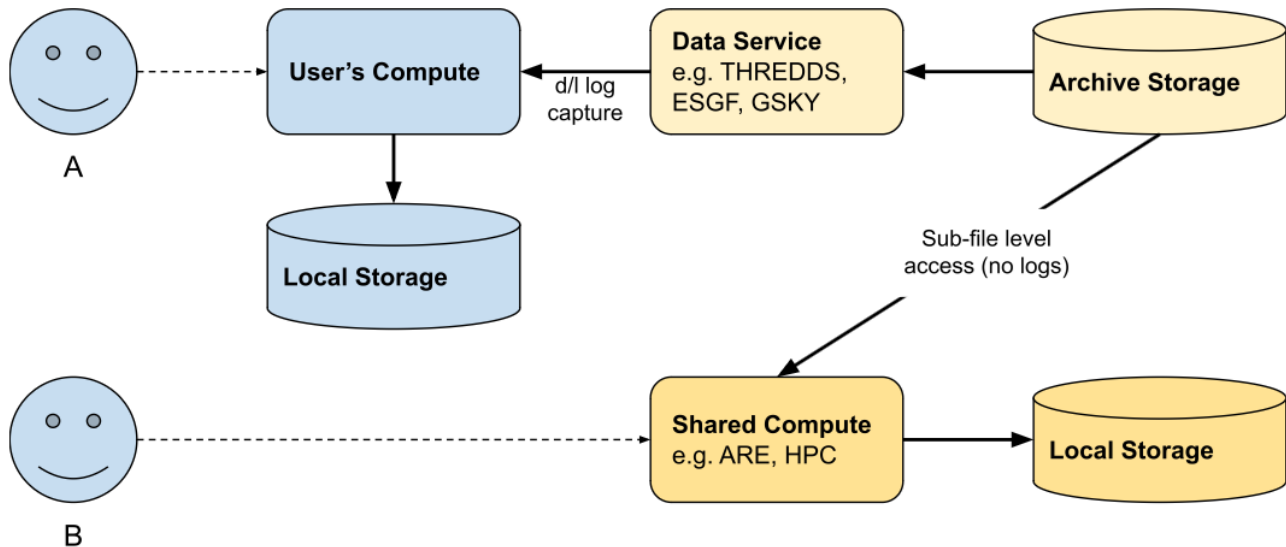
Multiple NCI Services/Platforms

The image displays three overlapping screenshots of NCI services:

- Top Left:** A search interface with a search bar containing "Search" and a "Back to search" button. Below is a snippet of text about "AusPass Passive Seism" and a description of the Passive Seismic Data.
- Top Center:** A project page for "Project qt52" under the "NCI AUSTRALIA" header. It features a navigation menu with "Projects and groups" selected. The main content area shows "Overview" and "Members" tabs, and a description of the project.
- Bottom Right:** A "Catalog" page for "NCI THREDDS Server" showing a "Dataset" section with "AusPass Passive Seismic Collection" and links to "license.txt" and "readme.txt". It also lists years "1993" and "1994" with folder icons.

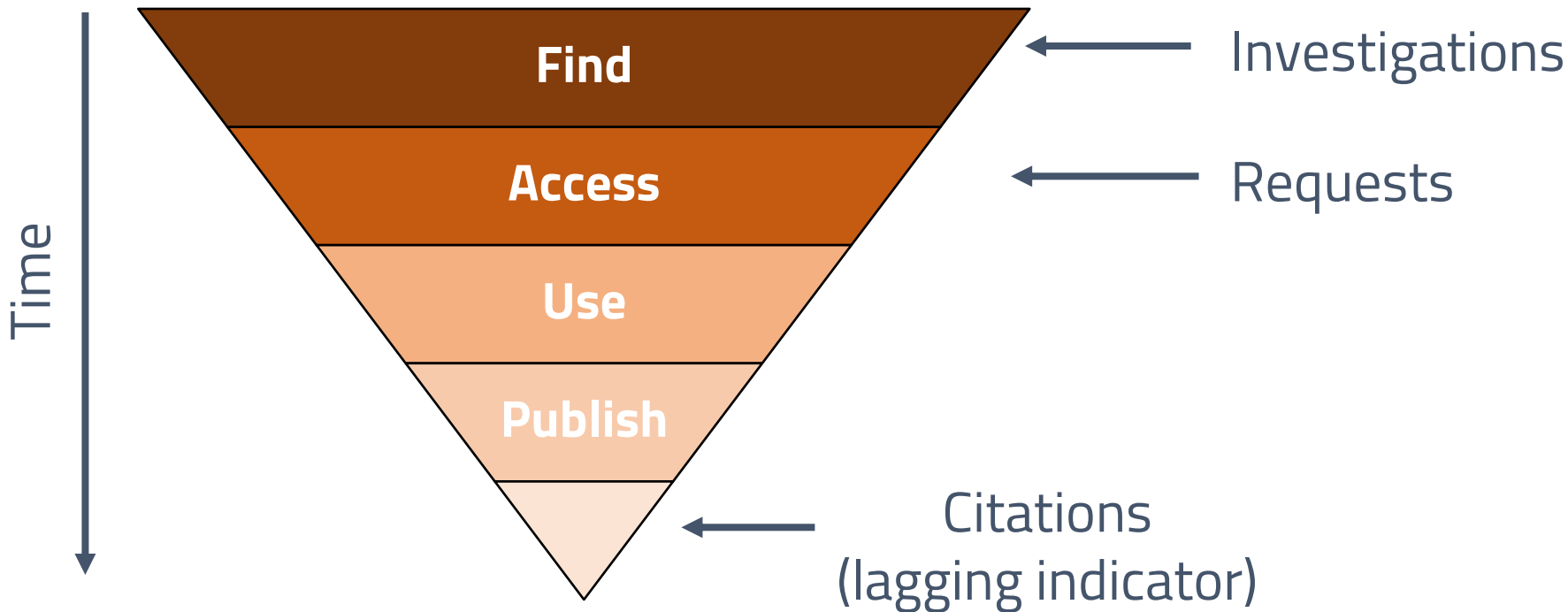
COUNTER & High Performance Computing

OPTION A: Download locally for analysis



OPTION B: Use NCI shared compute resources to directly read & perform analysis

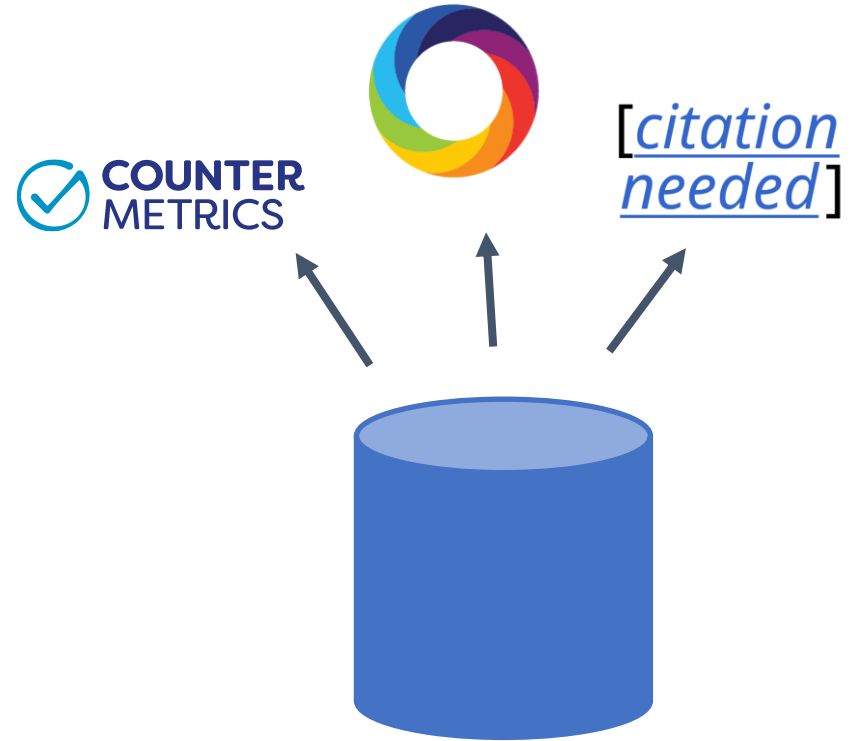
Usage tracking pipeline



Evidence-based data management

COUNTER RD metrics can be valuable for day-to-day data management:

- ✓ Monitor usage/demand from data users over time
- ✓ Help prioritise data for hosting/sharing (or decommissioning)



Applying metrics

*"All models are wrong,
but some are useful"*

*"No single usage metric is
perfect, however
standardised usage tracking
is potentially useful"*

NCI Australia. (2023). COUNTER Usage Metrics: National Computational Infrastructure - ARDC Project Report (1.0). NCI Australia.

<https://doi.org/10.5281/zenodo.8093111>



COUNTER Usage Metrics

National Computational Infrastructure - ARDC Project Report (v1.0)

EXECUTIVE SUMMARY

The National Computational Infrastructure (NCI) has completed an investigation into the implementation of standardised research data usage metrics as part of the ARDC-funded National Data Assets Project impact/outcome monitoring systems work package. The NCI THREDDIS data service was used as a case study for the implementation of COUNTER RD usage statistics. To better support usage statistics reporting in future, a number of design principles and NCI development priorities were identified. In addition, the project assessed the ability of the COUNTER RD standard to capture the complexity of data usage via HPC facilities.

BACKGROUND

The ARDC-funded National Data Assets Project included an impact/outcome monitoring systems work package. At NCI, this work package was focused on a trial implementation of the first release of the [COUNTER Code of Practice for Research Data Usage Metrics](#) (COUNTER RD). This standard provides a mechanism for capturing, processing, recording and sharing research data usage. The key metrics are the number of 'investigations' (i.e., access to the dataset metadata) and the number and volume of 'requests' (i.e., data downloads or data views/accesses), recorded per month for each dataset. COUNTER RD has been implemented by a number of high-profile international data repositories, including Dryad, Figshare and DataOne.

With capacity for over 150 petabytes of data storage, NCI provides a range of different high-performance data, storage, and computing services. NCI [data services](#) include an Earth Systems Grid Federation (ESGF) node, THREDDIS, GSKY, the Copernicus SARA Data Portal and various optical astronomy data services. NCI supports researchers' compute and data-intensive workloads. In order to access High Performance Computing (HPC) and Australian Research Environment (ARE), end-users require an NCI account and computing allocation. Usage reporting has previously focused on tracking data providers' storage limits and compute allocations for HPC end-users. The number of registered users is sometimes used as a proxy for the level of

1

Questions





NCI Contacts



General enquiries: +61 2 6125 9800



Support: help@nci.org.au



Website: nci.org.au



Address

NCI, ANU Building 143
143 Ward Road
The Australian National University
Canberra ACT 2601

License

