

Vertical Integration of National Geophysical Data Assets to Support Next-generation Reproducible Research at Exascale

eResearch Australasia 2023

Jo Croucher¹, Hannes Hollmann¹, Yiling Liu¹, Andrew Robinson¹, Nigel Rees¹, Rebecca Farrington², Lesley Wyborn¹, Ben Evans¹

¹ Australian National University, National Computational Infrastructure, Acton, Australia

² AuScope Ltd, University of Melbourne, Melbourne, Australia

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



Lynnice Letty Church (2020), *Gadi*, NCI Australia
Ngunnawal, Wiradjuri & Kamilaroi (ACT and NSW)



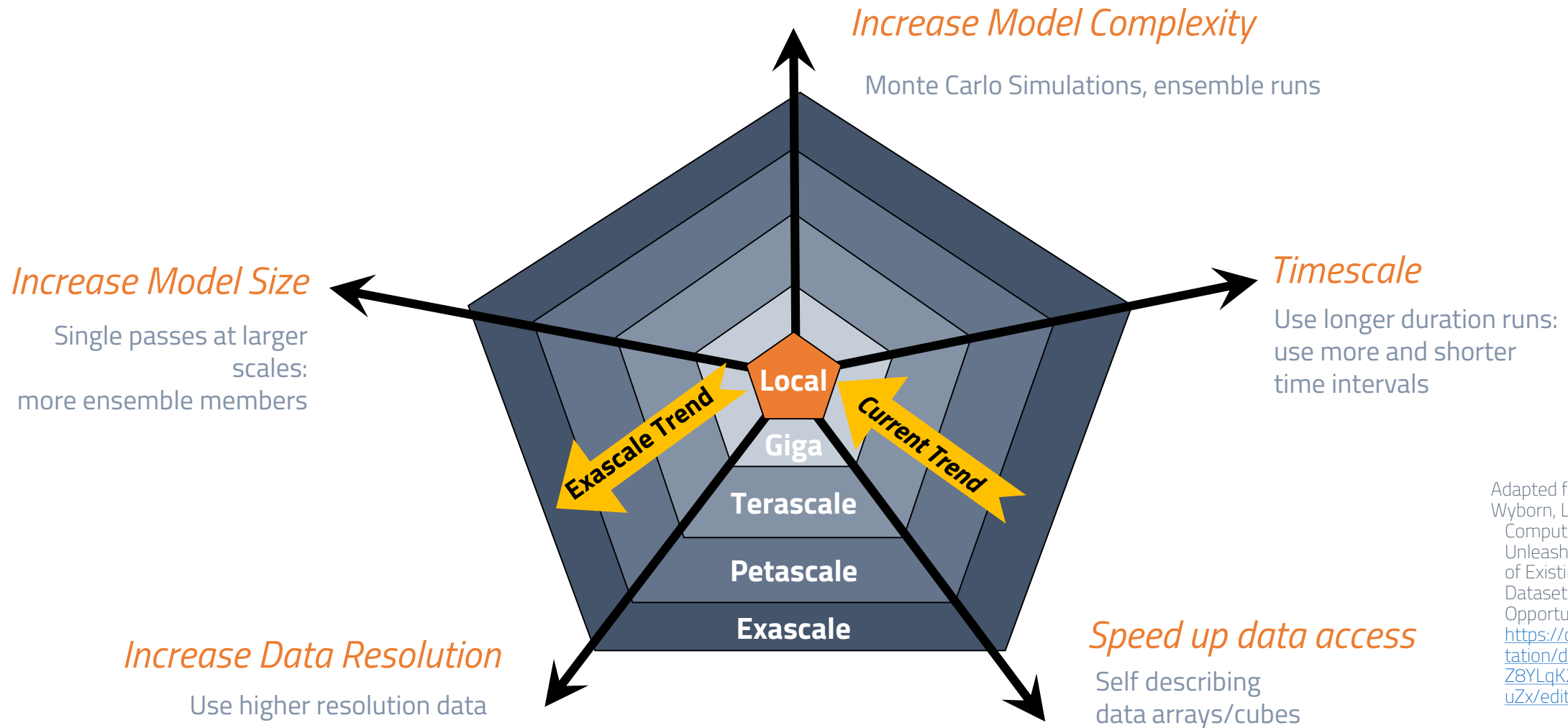
Next-generation geophysics

Geophysics at Exascale

- Exponential increase in data collected
- Massive decrease in processing times



Next-Generation Research



Adapted from:
Wyborn, L., 2023: Using 2030
Computational Techniques to
Unleash the Untapped Potential
of Existing Geophysical
Datasets in Mineral Exploration:
Opportunities and Challenges:
<https://docs.google.com/presentation/d/170lunofp5EpxNJ0iq8Z8YLqK2KvF-uZx/edit#slide=id.p1>

2030 Geophysics Collections

- Making national high-resolution geophysics reference collections suitable for exascale computing in 2030
- The project received co-investment from AuScope and the ARDC to support the data curation at NCI. AuScope, the ARDC and NCI are enabled by NCRIS.



4+
geophysics
collections
(160+ datasets)

30+ TiB
data



FAIR

Integrated

HPC-Enabled

Image: NASA

NCI Data Catalogue

Get Started

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

Browse by topics

- Climatology, meteorol... 283
- Geoscientific information 228
- Environment 154
- Oceans 50
- Biota 44
- Elevation 6
- Imagery base maps ea... 5
- Health 4
- Inland waters 1

Browse resources

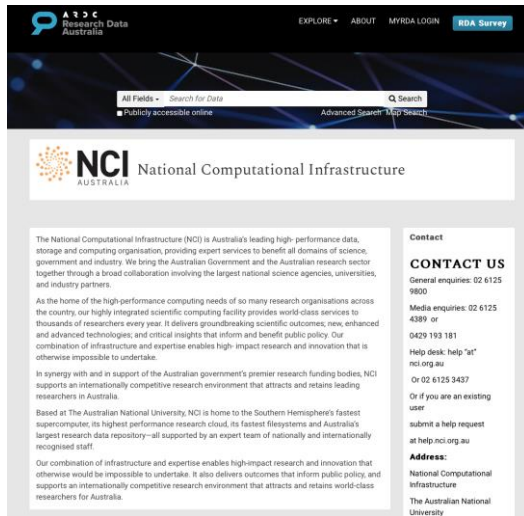
- Dataset 896
- Collection 91
- Service 8

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

Find Data

✓ Search the [NCI Data Catalogue](#) & navigate the collection

✓ Harvested by [Research Data Australia](#)



 Magnetotelluric surveys of the AusLAMP Musgraves Province within Western Australia, 2016 to 2018 (Level 0 concatenated ASCII time series per run)

As part of the Australian Lithospheric Architecture Magnetotelluric Project (AusLAMP), long-period magnetotelluric time series data were collected in the Musgraves Province between 2016 and 2018. Level 0 concatenated ASCII time series data from Earth Data Logger sites located in Western Australia are presented per run.

Data Access

Register for local access <https://my.nci.org.au/mancini/project/my80/join>

NCI local file path
/g/data/my80/AuScope_MT_collection/AuScope_AusLAMP
/Musgraves_APY/WA/level_0
/Concatenated_Time_Series_ASCII_per_run



NCI THREDDS Data Server https://dapds00.nci.org.au/thredds/catalog/my80/AuScope_MT_collection/AuScope_AusLAMP/Musgraves_APY/WA/level_0/Concatenated_Time_Series_ASCII_per_run/catalog.html

Open link



Foundation Paper

Thiel, S., Goleby, B.R., Pawley, M.J. et al. (2020). AusLAMP 3D MT imaging of an intracontinental deformation zone, Musgrave Province, Central Australia. Earth Planets Space, 72, 98. <https://doi.org/10.1186/s40623-020-01223-0>

Open link

Access & Reuse Data

Data Access

Register for local access

<https://my.nci.org.au/mancini/project/my80/join>

NCI local file path

/g/data/my80/AuScope_MT_collection/AuScope_AusLAMP
/Musgraves_APY/SA/level_0
/Concatenated_Time_Series_ASCII_per_run/



NCI THREDDS Data Server https://dapds00.nci.org.au/thredds/catalog/my80/AuScope_MT_collection/AuScope_AusLAMP/Musgraves_APY/SA/level_0/Concatenated_Time_Series_ASCII_per_run/catalog.html

Open link

← Access

Access constraints

<https://creativecommons.org/licenses/by/4.0/>
Creative Commons Attribution 4.0 International

← Check

Catalogue record DOI

10.25914/y380-ds12
URL: [<https://dx.doi.org/10.25914/y380-ds12>]

← Cite

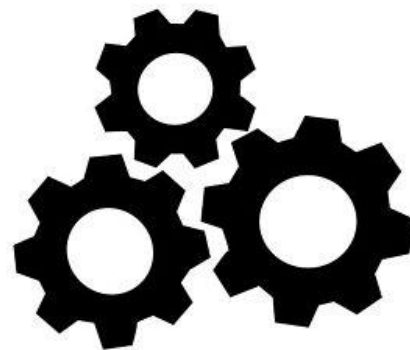
F
Findable



A
Accessible



I
Interoperable



R
Reusable

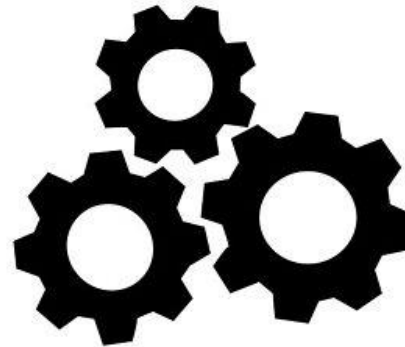


F
Findable

A
Accessible

I
Interoperable

R
Reusable



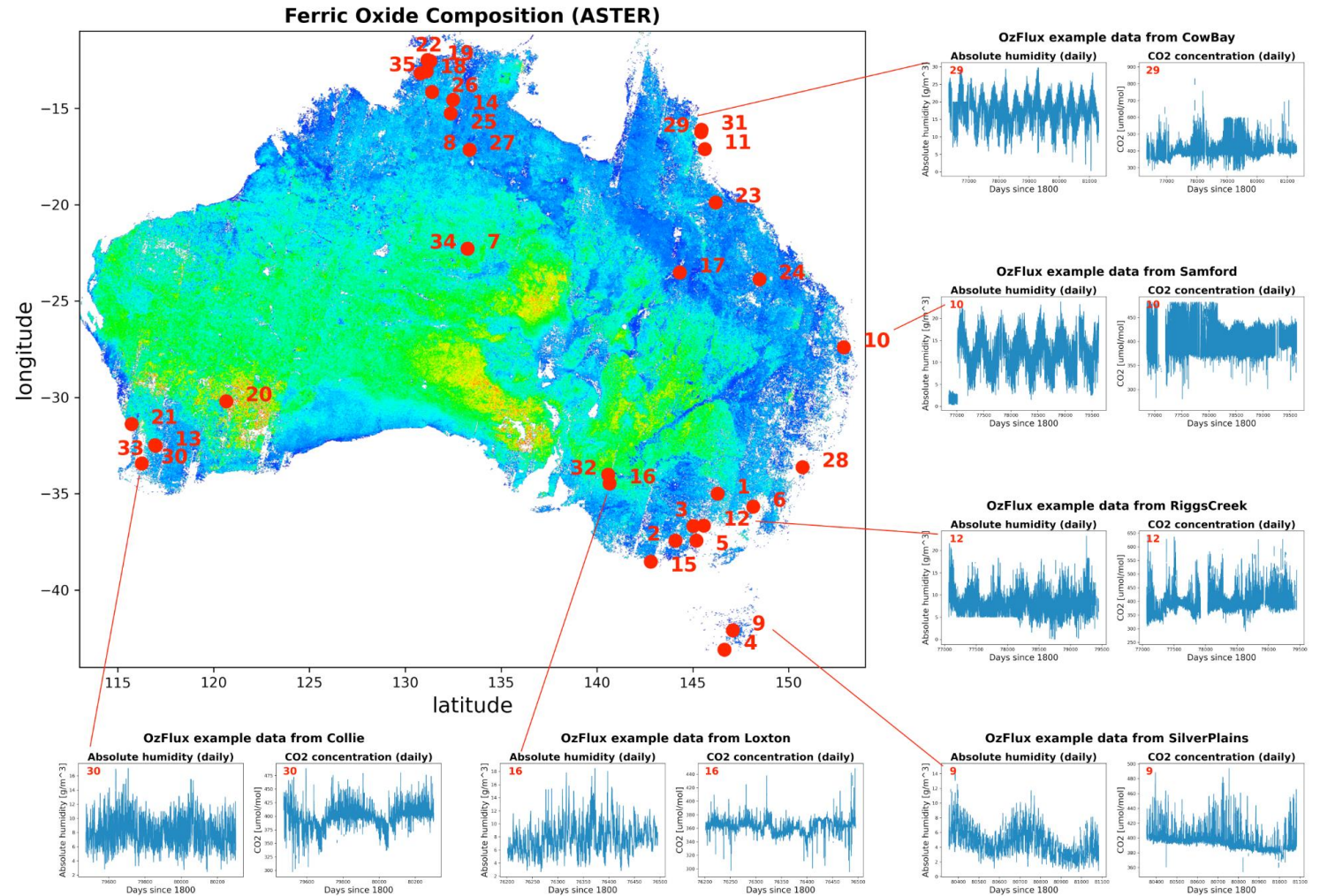
...The FAIR Principles put specific emphasis on enhancing the ability of **machines** to **automatically find and use the data**, in addition to supporting its reuse by individuals.

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

Image: [File:Default-welcomer.png](#)

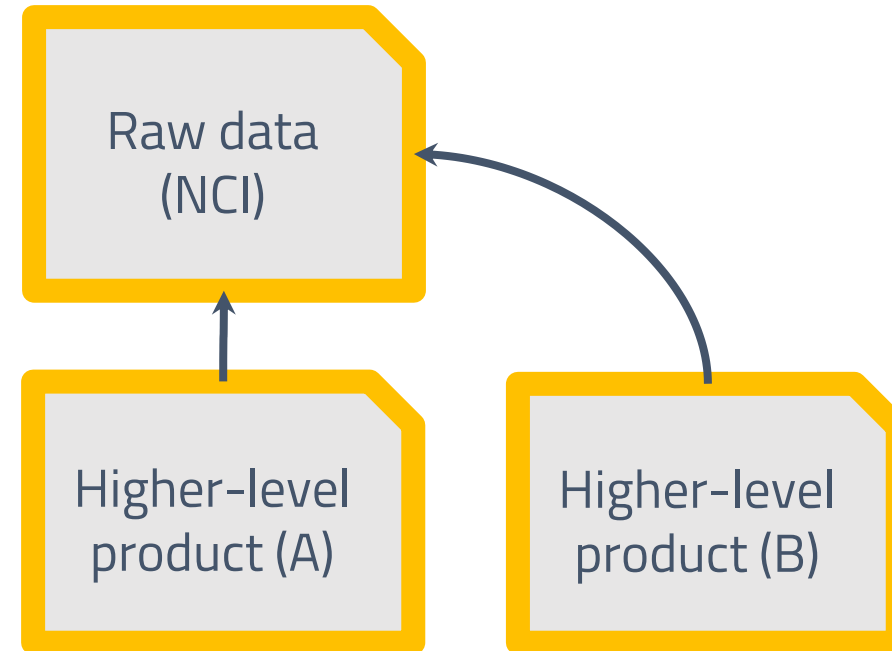
Horizontal Integration of Data from External Repositories

- Integrating two disparate datasets to create a new derivative product
- For example, external data (TERN's OzFlux tower locations (red dots)) is integrated with data hosted at NCI (ASTER layer)



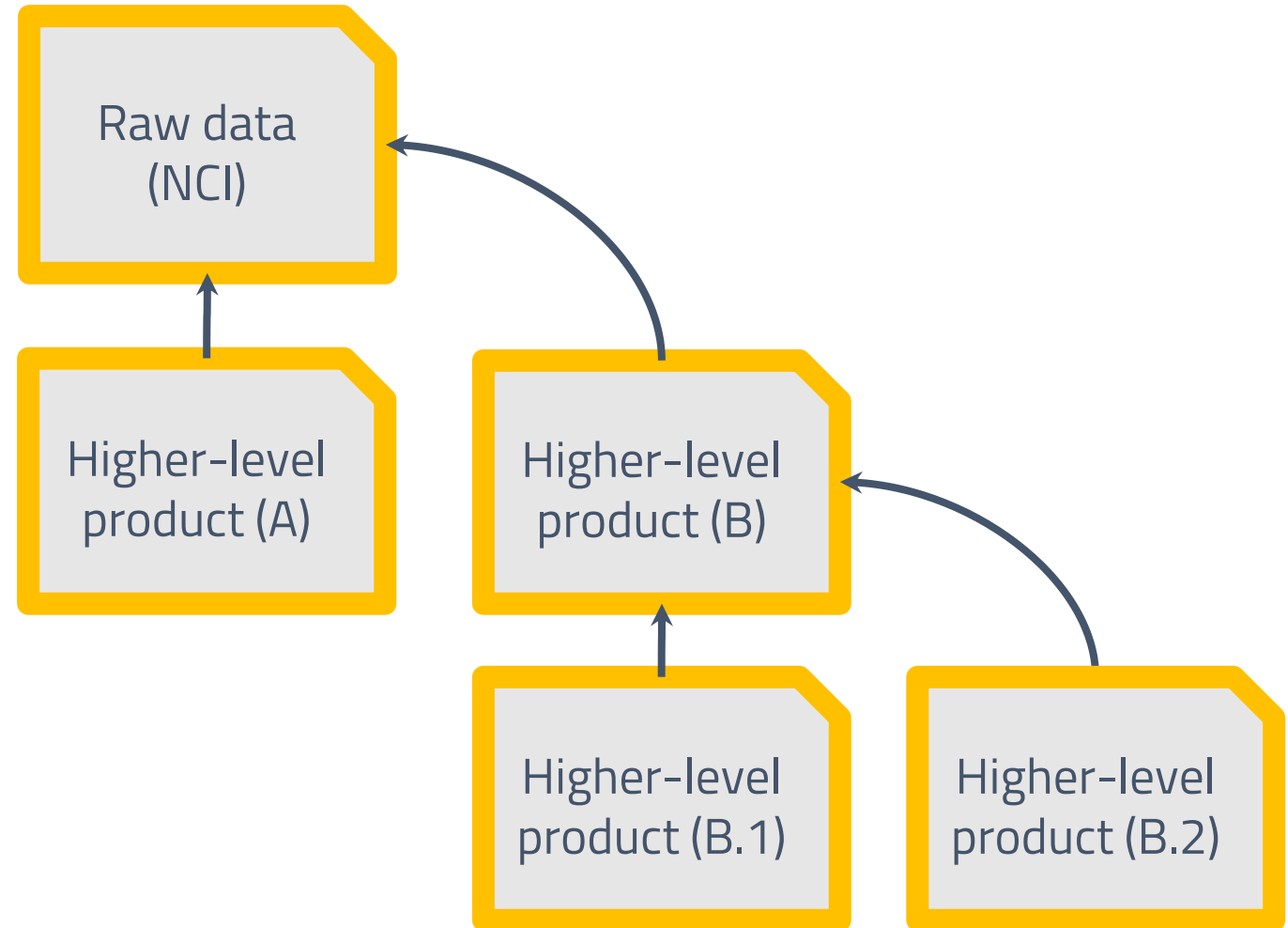
Vertical Integration

- Catalogue metadata is structured to enable 'vertical' integration between repositories
- Higher-level products (from various sources) reference the raw data hosted at NCI

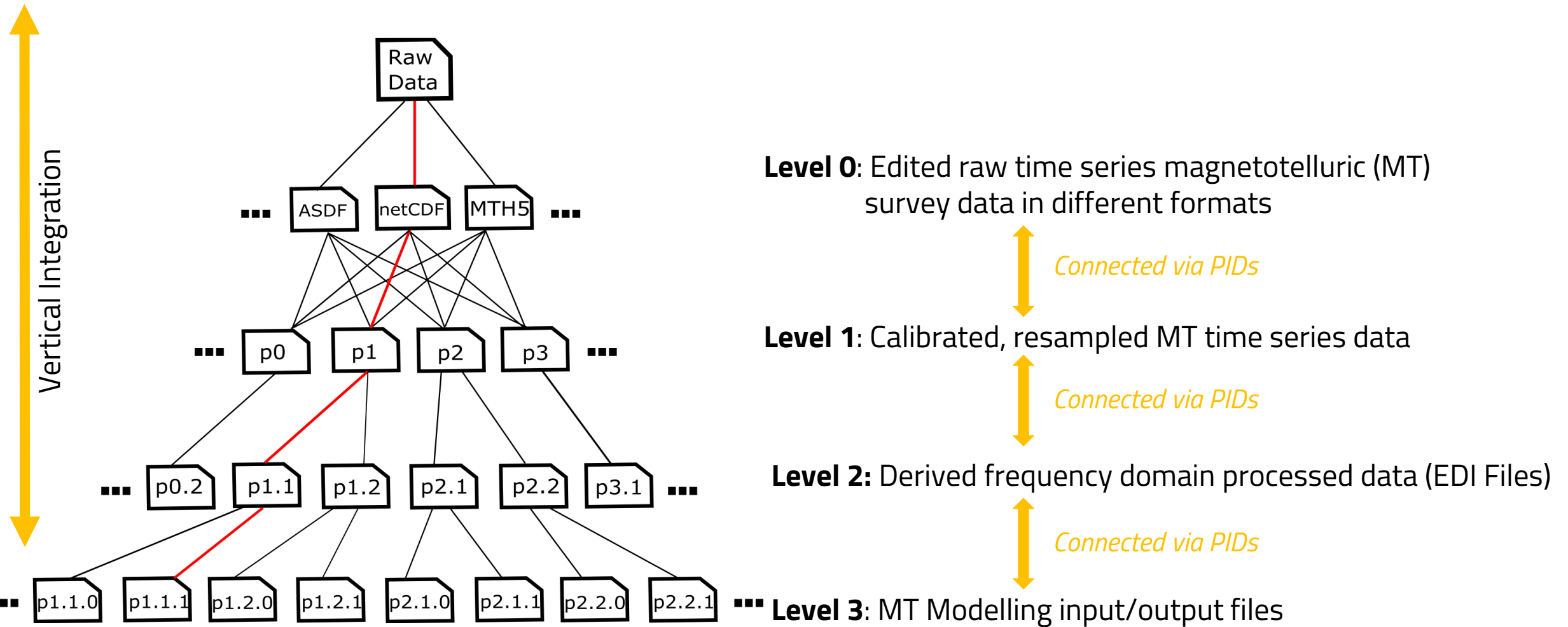


Vertical Integration

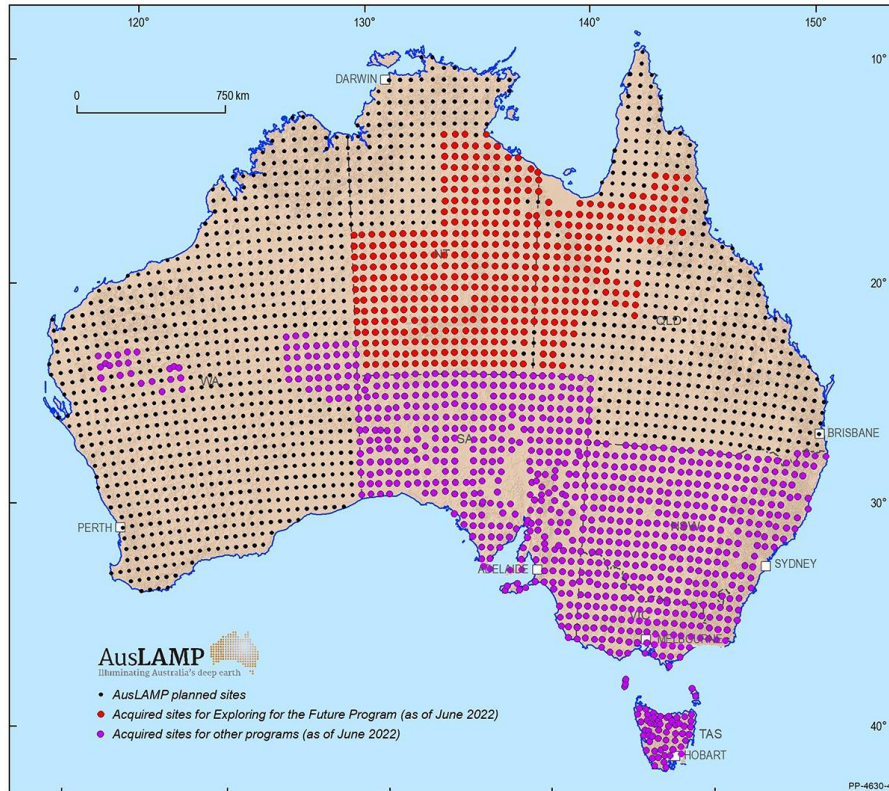
- Catalogue metadata is structured to enable 'vertical' integration between repositories
- Higher-level products (from various sources) reference the raw data hosted at NCI



Vertical Integration



AusLamp



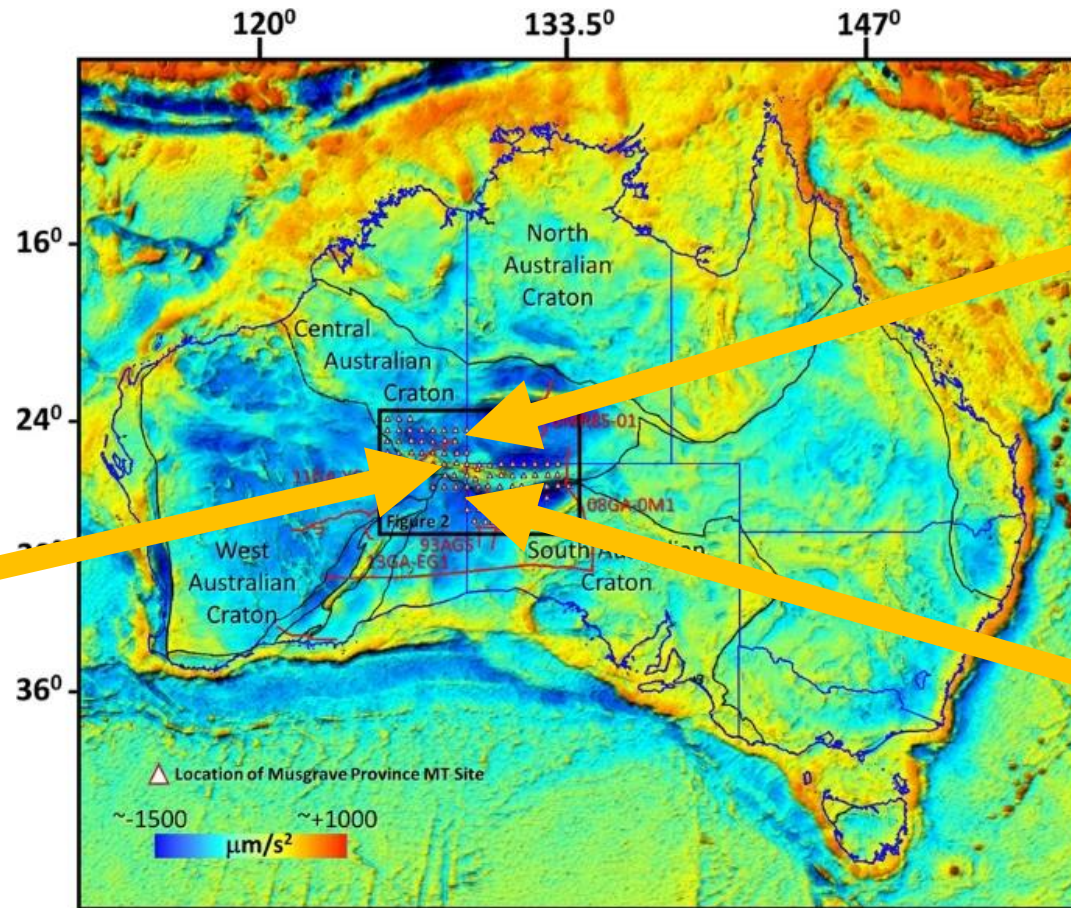
© Commonwealth of Australia
(Geoscience Australia) 2022



© Commonwealth of Australia
(Geoscience Australia) 2021

<https://www.ga.gov.au/about/projects/resources/auslamp>

Musgraves Province



WA = CC-BY 4.0

NT = CC-BY 4.0

SA = CC-BY 3.0

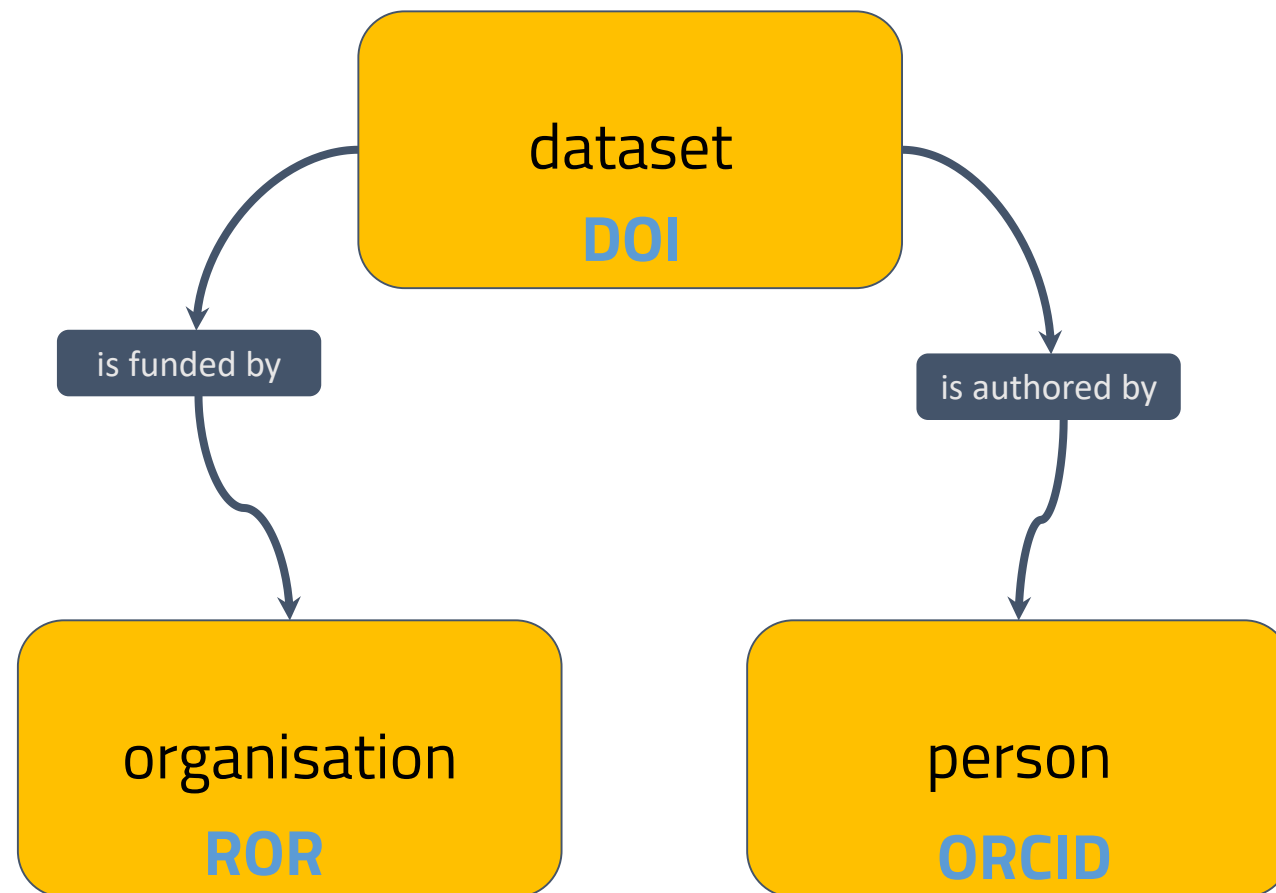
Thiel, S., Goleby, B.R., Pawley, M.J. et al. AusLAMP 3D MT imaging of an intracontinental deformation zone, Musgrave Province, Central Australia. *Earth Planets Space* 72, 98 (2020). <https://doi.org/10.1186/s40623-020-01223-0>

Vertical Integration: Persistent Identifiers

Persistent identifiers promote FAIR by:

✓ Enabling citation and attribution

✓ Improved machine-readability



Lineage Statements

✓ Include DOI for raw source data

✓ Document processing steps

✓ Include DOI for software used to process the data

Lineage

This dataset was derived from a raw source dataset first published by the University of Adelaide (Heinson, G. 2022).

The MagnetoTellurics time series data publication (MTtsdp) codes (Rees, N. et al. 2023) were used to create the Level 1 concatenated, resampled, rotated MTH5/MT_metadata time series data products. The Level 0 concatenated MTH5/MT_metadata time series data products (<https://dx.doi.org/10.25914/9y8g-9x90>) were used as input for the generation of the Level 1 concatenated, resampled, rotated MTH5/MT_metadata time series data products.

References:

Heinson, G. (2022). Magnetotelluric surveys of the AusLAMP Musgraves Province within Western Australia, 2016 to 2018 (packed raw time series data) [Data set]. NCI Australia.

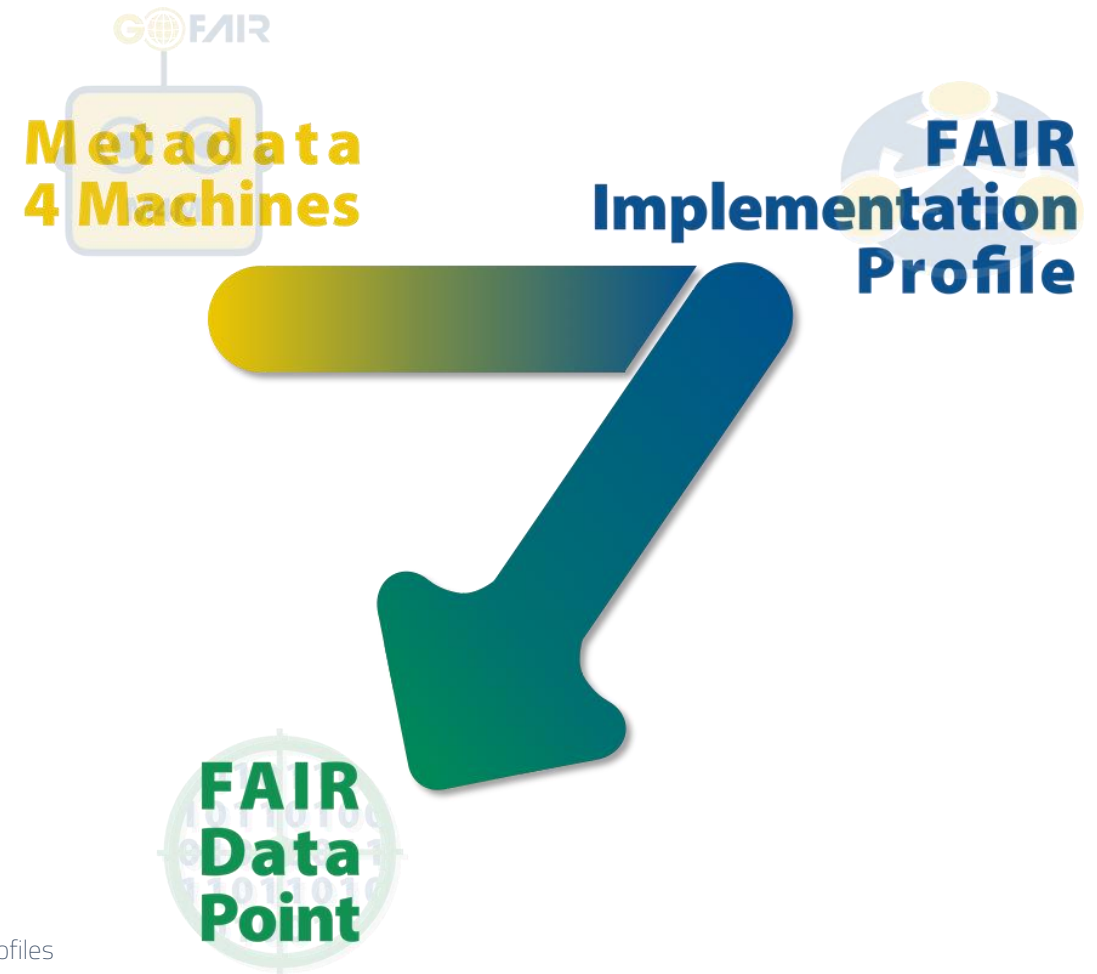
<https://doi.org/10.25914/AJ1K-FJ69>

Rees, N., Wang, S., Conway, D. et. al. (2023). nci/MTtsdp: MTtsdp_v1.0.0 (v1.0.0) [Software]. Zenodo.

<https://doi.org/10.5281/zenodo.7707364>

FAIR Implementation Profile (FIP)

- Document FAIR implementation choices
- For each FAIR Principle, provides a link to the exact resources used
- Promotes interoperability and standardisation
- Designed to be machine-actionable



Schultes, E., Magagna, B., Hettne, K.M., Pergl, R., Suchánek, M., Kuhn, T. (2020). Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. In: Grossmann, G., Ram, S. (eds) Advances in Conceptual Modeling. ER 2020. Lecture Notes in Computer Science, vol 12584. Springer, Cham. https://doi.org/10.1007/978-3-030-65847-2_13

Build your FAIR Implementation Profile

This is a template. Please copy the file before filling it out.

Community description	
Name of Community	e.g. ENVRI
Description of Community	
Supporting Links	
Research Domain	e.g. Environmental Sciences
Data Steward	e.g. ORCID #
Date of FIP creation	

→ PDF

FAIR principle	Question	FAIR enabling resource types	Your answers
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type	e.g. PURL, DOI
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type	
F2	Which metadata schemas do you use for findability?	Metadata schema	
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism	
F4	In which search engines are your metadata records indexed?	Search engines	
F4	In which search engines are your datasets indexed?	Search engines	
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol	
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol	
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique	
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique	
A2	Which metadata longevity plan do you use?	Metadata longevity	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language	
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies	
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies	
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema	
I3	Which models, schema(s) do you use for your datasets?	Data schema	
R1.1	Which usage license do you use for your metadata records?	Data usage license	
R1.1	Which usage license do you use for your datasets?	Data usage license	
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	

→ Excel

GoFAIR (2023), FAIR Implementation Profile, <https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/>

FIP Wizard

AuScope Magnetotellurics (MT) Collection Share

Questionnaire Metrics Preview Documents Settings

View TODOs Comments Version history

Current Phase

Defining FAIR Implementation Profile

Chapters

- I. About ✓
- II. Declare your FAIR Implementation Community ✓
- III. Declarations for Findability ✓
- IV. Declarations for Accessibility ✓
- V. Declarations for Interoperability ✓
- VI. Declarations for Reusability ✓
- VII. Register a new resource as a nanopublication ✓

I. About

A FAIR Implementation Profile (FIP) is a list of declared technology choices intended to implement each of the FAIR Guiding Principles, made as a collective decision by the members of a particular community of practice. As can be seen herein, the FIP Wizard captures FIPs by means of a questionnaire that prompts a representative of the community (the Community Data Steward) to provide answers that explicitly profile the FAIR implementation approach of that community. FIPs are published by the FIP Wizard as FAIR (machine-readable) and Open data, which can then serve as a reference for practical FAIR data stewardship activities conducted by members of that community. FIP publication also encourages FIP reuse and repurposing by other communities, which saves time 'reinventing the wheel' and simultaneously drives convergence on FAIR implementation choices. Over time, FIPs will need to be revised to reflect the evolving needs of the community and the ongoing development of FAIR technologies. In short, the FIP Wizard helps to make FAIRification efforts more structured, better informed and overall more efficient.

The FIP Wizard has a number of features specifically supporting FIP creation:

Questionnaire: The FIP questionnaire is augmented with explanations, per question, based on Jacobsen et al. FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence 2020; 2 (1-2): 10–29, DOI and referencable at the GO FAIR Foundation website

Navigation: While creating the FIP, the questionnaire can be navigated using the navigation tool on the left side of the page.

Versioning: The FIP Wizard has versioning that allows systematic revisions of completed FIPs over time which in turn can provide insight into FAIR-related technology trends.

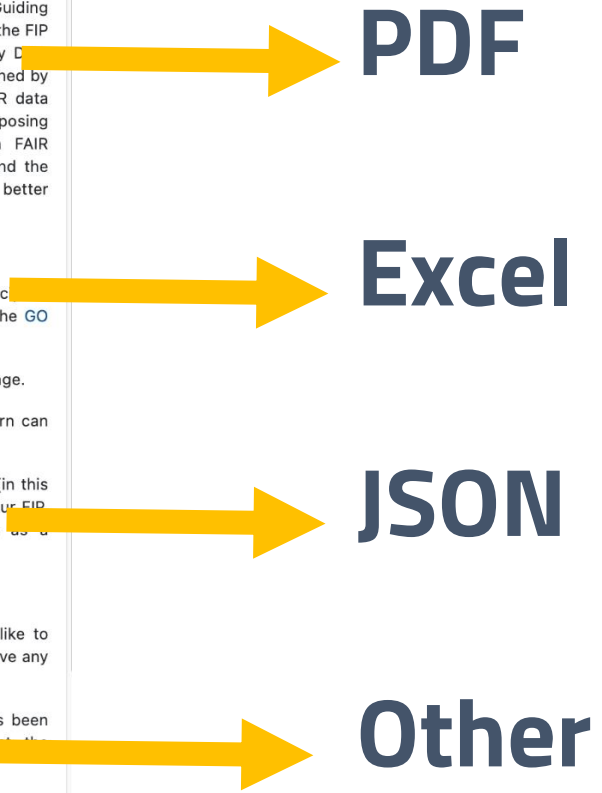
Nanopublications: The FIP Wizard makes use of a data format called nanopublications to capture FIPs as FAIR data (in this case, as FAIR Digital Objects). In some cases, it may be necessary to author original nanopublications to complete your FIP. The FIP Wizard supports the creation of original nanopublications in Chapter VII, "Register a new resource as a nanopublication".

Detailed instructions for completing the FIP Wizard questionnaire can be found [here](#).

FIP Wizard team is composed of Barbara Magagna, Marek Suchanek, Tobias Kuhn, and Erik Schultes. We would like to acknowledge the generous support of CODATA and ENVRI-FAIR. We wish you success in composing your FIP. If you have any comments, questions or suggestions on how to improve the FIP Wizard experience, [contact us](#).

The FIP Wizard has its origin in the [GO FAIR FAIR Convergence Matrix \(& FIP\) Working Group](#) and development has been supported by the Center for Digital Scholarship at the Leiden University Libraries, CODATA, the ENVRI-FAIR project and the enthusiastic participation of numerous FAIR Implementation Communities, see.

For further reading: Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence. In: Grossmann G., Ram S. (eds) Advances in Conceptual Modeling. ER 2020. Lecture Notes in Computer Science, vol 12584. Springer, Cham, DOI.



FIP Wizard (2023): <https://fip-wizard.ds-wizard.org/>

NCI Zenodo Community

- ✓ Publish FIPs (in multiple file formats)
- ✓ Generate DOIs
- ✓ Manage versions

zenodo Search Upload Communities Log in Sign up

National Computational Infrastructure - FAIR Implementation Profiles

Recent uploads

Search National Computational Infrastructure - FAIR Implementation Profiles

July 28, 2023 (1.0) Other Open Access View

NCI Data Catalogue - FAIR Implementation Profile (FIP)

NCI Australia;

This FAIR Implementation Profile (FIP) describes the NCI Data Catalogue (<https://geonetwork.nci.org.au/>) published by the National Computational Infrastructure (NCI). It outlines the FAIR implementation choices implemented for each of the FAIR Principles. The FIP was prepared via

Uploaded on August 25, 2023

July 28, 2023 (1.0) Other Open Access View

AuScope Distributed Acoustic Sensing (DAS) Collection - FAIR Implementation Profile (FIP)

NCI Australia;

This FAIR Implementation Profile (FIP) describes the AuScope Distributed Acoustic Sensing (DAS) Collection published by the National Computational Infrastructure (NCI). It outlines the FAIR implementation choices implemented for each of the FAIR Principles. The FIP was prepared via

Uploaded on August 25, 2023

July 28, 2023 (1.0) Other Open Access View

AusPass Passive Seismic Collection - FAIR Implementation Profile (FIP)

NCI Australia;

This FAIR Implementation Profile (FIP) describes the AusPass Passive Seismic Collection published by the National Computational Infrastructure (NCI). It outlines the FAIR implementation choices implemented for each of the FAIR Principles. The FIP was prepared via the FI

New upload

Community

National Computational Infrastructure - FAIR Implementation Profiles

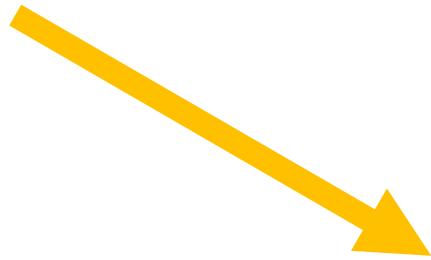
The [National Computational Infrastructure \(NCI\)](https://nci.org.au/) is Australia's leading high-performance data, storage and computing organisation, providing expert services to benefit all domains of science, government and industry.

This community hosts NCI's FAIR Implementation Profiles (FIPs). These resources document the FAIR implementation practices of datasets and collections hosted at NCI. Each FIP provides machine-readable information regarding implementation of the [FAIR Principles](https://www.fairprincipl.es/) (i.e. Findable, Accessible, Interoperable, Reusable).

More information about NCI's services, training and events is available from the NCI Australia website: <https://nci.org.au/>

Integrating FIPs

✓ Published FIPs can be cited and linked to relevant resources



AuScope Magnetotellurics (MT) Collection

This collection includes AuScope-funded geophysical data, as well as data that has been acquired by universities, industry, federal/state government agencies since the 1950s. Magnetotelluric (MT) survey data has been collected from thousands of sites across Australia.

Data Access

Register for local access <https://my.nci.org.au/mancini/project/my80/join>

NCI local file path /g/data/my80/AuScope_MT_collection/



NCI THREDDS Data Server

<https://dapds00.nci.org.au/thredds/catalogs/my80/catalog.html>

Open link



FAIR Implementation Profile (FIP)

<https://doi.org/10.5281/zenodo.8185349>

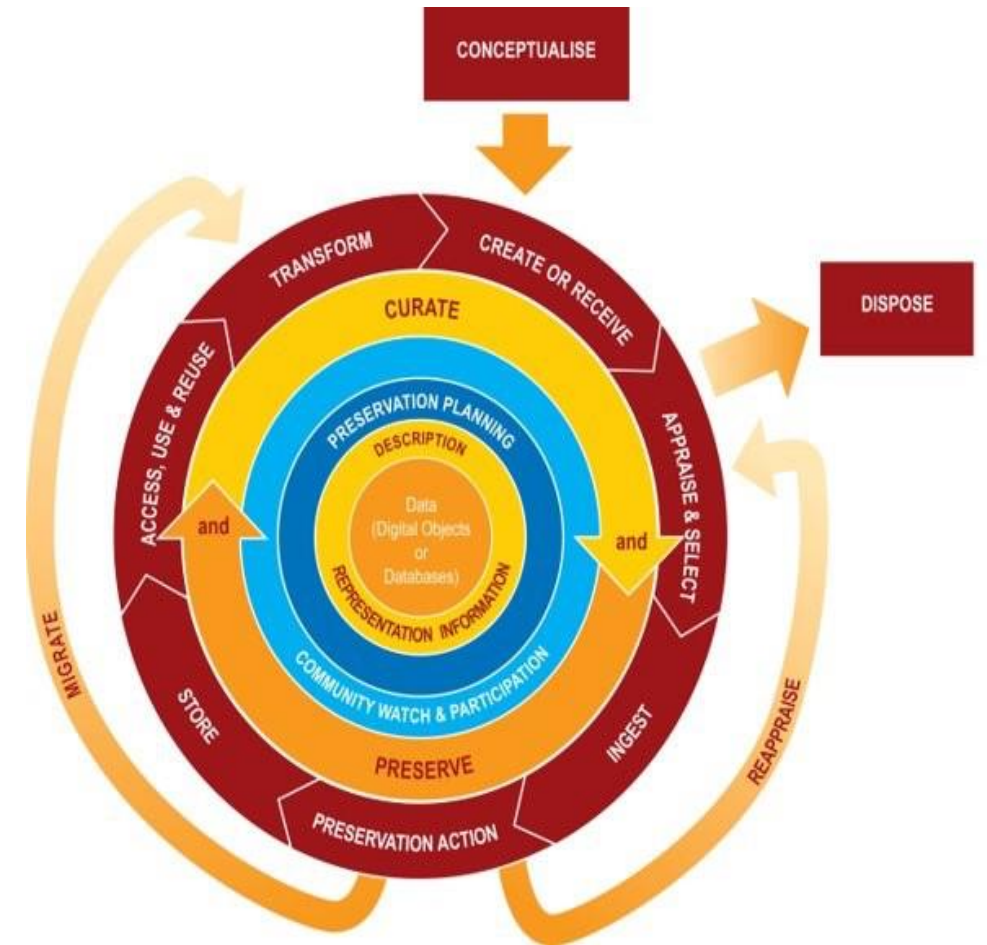
Open link



Next-generation geophysics

Reusable

- ✓ Plan for reuse up-front
- ✓ Ensure infrastructure designed to support FAIR and CARE principles
- ✓ Consider range of approaches to measuring impact



DCC (2004-2023), The DCC Curation Lifecycle Model:
<https://doi.org/10.2218/ijdc.v3i1.48>

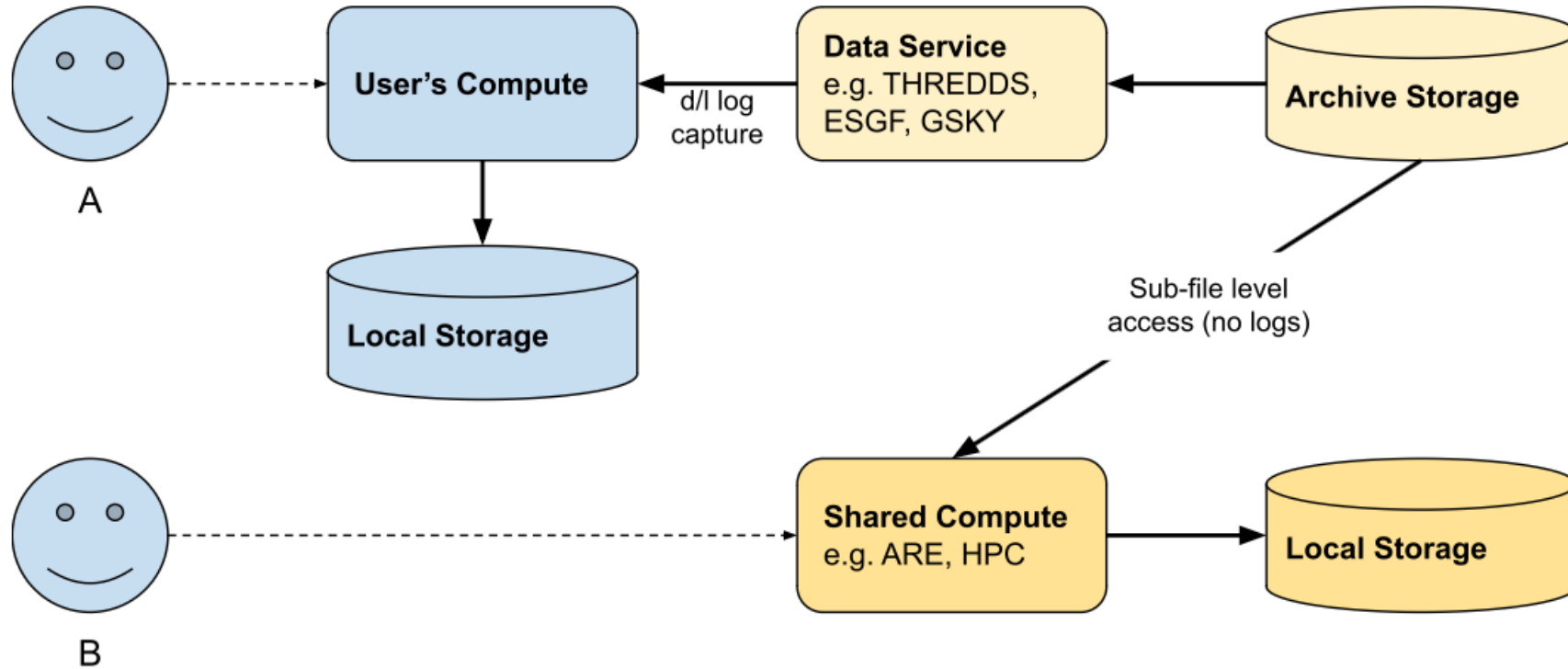
Tracking Use and Reuse

The COUNTER Code of Practice for Research Data

- ✓ Standardised usage statistics for research data
- ✓ Tracks “investigations” (i.e., metadata views)
- ✓ Tracks number and volume of “requests” (i.e., data downloads or data views/access)



COUNTER RD & HPC



OPTION A: Download locally for analysis

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

Preparing for 2030...

- ✓ Provide access to the raw high-resolution observational data
- ✓ Invest in data management and curation (e.g. automated 'dirt-to-desktop' workflows)
- ✓ Consider plans for machine-interoperability (e.g. PIDs, FIPs)



Questions





NCI Contacts



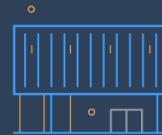
General enquiries: +61 2 6125 9800



Support: help@nci.org.au



Website: <https://opus.nci.org.au/x/NAOVAg>



Address

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

License

