NCI AUSTRALIA

Preparing geophysical data and software for Exascale: A case study from the 2030 Geophysics Collection Project

eResearch Australasia 2023

Nigel Rees, Rui Yang, Yue Sun, Edison Guo, Lesley Wyborn, Ben Evans

19 October 2023

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

Artist: Lynnice Letty Church – Tribes: Ngunnawal, Wiradjuri & Kamilaroi (ACT and NSW)
Gadi - "to search for" in Ngunnawal language - January 2020 for NCI Gadi Supercomputer

The 2030 Geophysics Collection Project

- 2030 was an R&D project funded through a collaboration between AuScope, NCI, TERN and ARDC (https://ardc.edu.au/project/2030-geophysics-collections/)

- It seeked to:
  a. Make national-scale high-resolution geophysics datasets suitable for programmatic access in HPC environments;
  b. Lay the foundations for more rapid data processing by 2030 next-generation scalable, data intensive computation including Artificial Intelligence (AI) / Machine Learning (ML) and data assimilation.

- The project was about positioning Australian geophysical data collections to be capable of taking advantage of next generation technologies and computational infrastructures by 2030.

NCI AUSTRALIA

What do we know about 2030 computing?

1. High-end computational power will be at exascale
2. Today's emerging collaborative platforms will continue to evolve as a mix of HPC and cloud
3. Data volumes will be measured in Zettabytes (10^{21} bytes), which is about 10 times more than today
4. **It will be mandatory for data discovery, accessibility, interoperability and reusability to be fully machine-to-machine as envisaged by the FAIR principles in 2016**

Exascale computing is already a reality in the USA

In April 2018, the U.S. Department of Energy announced its intentions to procure a trio of exascale supercomputers at a cost of up to USD$1.8 billion.

Exascale computing is already a reality in the USA

In April 2018, the U.S. Department of Energy announced its intentions to procure a trio of exascale supercomputers at a cost of up to USD$1.8 billion.

In 2022, the HPE **Frontier** system was released as the world's first exascale supercomputer, hosted at the Oak Ridge Leadership Computing Facility.

https://en.wikipedia.org/wiki/Frontier_%28supercomputer%29

Exascale computing is already a reality in the USA

In April 2018, the U.S. Department of Energy announced its intentions to procure a trio of exascale supercomputers at a cost of up to USD$1.8 billion.

In 2022, the HPE **Frontier** system was released as the world's first exascale supercomputer, hosted at the Oak Ridge Leadership Computing Facility.

https://en.wikipedia.org/wiki/Frontier_%28supercomputer%29

The Argonne National Laboratory's **Aurora** system is almost operational and is expected to be the world's first supercomputer to achieve a theoretical peak performance of more than 2 exaFLOPS.

https://en.wikipedia.org/wiki/Aurora_(supercomputer)

Exascale computing is already a reality in the USA

In April 2018, the U.S. Department of Energy announced its intentions to procure a trio of exascale supercomputers at a cost of up to USD$1.8 billion.

In 2022, the HPE **Frontier** system was released as the world's first exascale supercomputer, hosted at the Oak Ridge Leadership Computing Facility.

https://en.wikipedia.org/wiki/Frontier_%28supercomputer%29

The Argonne National Laboratory's **Aurora** system is almost operational and is expected to be the world's first supercomputer to achieve a theoretical peak performance of more than 2 exaFLOPS.

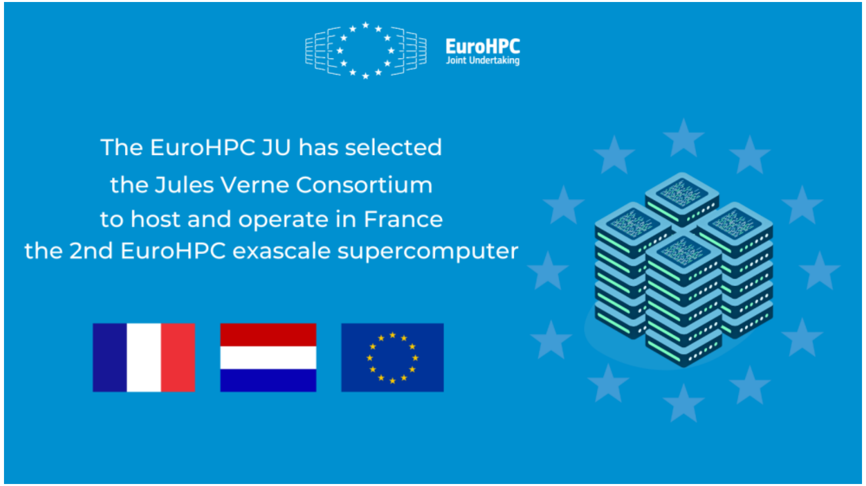
https://en.wikipedia.org/wiki/Aurora_(supercomputer)

The Lawrence Livermore National Laboratory's **El Capitan** system is set to come online in mid-2024 and is expected to deliver performance of over 2 exaFLOPS.

https://www.anandtech.com/show/15581/el-capitan-supercomputer-detailed-amd-cpus-gpus-2-exaflops

Exascale machines will soon arrive in Europe

https://www.fz-juelich.de/en/ias/jsc/jupiter



https://eurohpc-ju.europa.eu/jules-verne-consortium-will-host-new-eurohpc-exascale-supercomputer-france-2023-06-20_en

Exascale machines will soon arrive in Europe .... and the UK

SPRING BUDGET 2023

3.92 In line with two of the key recommendations of the Future of Compute Review, the government will invest, subject to the usual business case processes, in the region of £900 million to build an exascale supercomputer and to establish a new AI Research Resource, with initial investments starting this year.[116] Together, these will provide significant compute capacity to our AI community and provide scientists with access to cutting-edge computing power. They will allow researchers to better understand climate change, power the discovery of new drugs and maximise our potential in AI.

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1144442/Print_Budget_2023.pdf

Edinburgh to lead new era of UK supercomputing

The UK's first next-generation supercomputer – 50 times faster than any of the country's existing machines – is to be hosted by the University of Edinburgh.

A new £31m wing of the Advanced Computing Facility has been built for the exascale supercomputer

https://www.ed.ac.uk/news/2023/edinburgh-to-lead-new-era-of-uk-supercomputing

What are the opportunities of 2030 computing?

So often today's research is undertaken on pre-canned, analysis-ready datasets (ARD) that are tuned towards the highest common denominator as determined by the data owner/publisher.

By 2030:

a. Increased computational power co-located with fast-access storage systems will mean that geophysicists will be able to work on less processed data levels and then transparently develop their own derivative products.
b. Researchers will be able to see the quality of their algorithms more rapidly: there will be multiple versions of open source software used as researchers fine tune individual algorithms to suit their specific requirements.
c. We will be capable of more precise solutions and in hazards space and other relevant areas, analytics will be done in faster-than-real-time.

Wyborn, L., Rees, N., Klump, J., Evans, B., Rawling, T., and Druken, K.: The Known Knowns, the Known Unknowns and the Unknown Unknowns of Geophysics Data Processing in 2030 , EGU General Assembly 2022, Vienna, Austria, 23–27 May 2022, EGU22-11012, https://doi.org/10.5194/egusphere-egu22-11012 , 2022.

Landsat-8 image courtesy of the U.S. Geological Survey

Expanding our geophysical data collections on HPC

Geoscience Australia Geophysics Reference Data Collection

This collection has been compiled by Geoscience Australia from an extensive archive of over 2,200 geophysical surveys dating back to 1947. The datasets have been acquired by Geoscience Australia and its State and Territory Government partners. The collection includes datasets acquired using potential field
(magnetics and gravity) and ra
elevation datasets (digital elev
geophysical surveys, are also
on the individual surveys for e
radiometric and elevation surv
the national scale compilation
and are stored in both their fu
interpolated regular grid (raste
disciplines including geologica
environmental resource manag

AusPass Passive Seismic Collection

The Passive Seismic Data contains mirrored datasets hosted on the Australian Passive Seismic Server (AusPass). AusPass is an initiative supported by funding from AuScope and the Australian National University to host passive seismic (i.e. continuous waveform) data via the standardised FDSN Web Service
protocol. Data largely consi
individual stations, but also
network and mirrors compli
Seismic Network (IU), amor
in individual network metad
AusPass website and data

References:
Albuquerque Seismological
[Data set]. International Fed
Australian Research Data C
geophysics reference colle
Balfour, N.J., Salmon, M. &
Education, outreach, resear
https://doi.org/10.1785/022
Geoscience Australia. (2021
September 2018). Commor
Salmon, M., Pickle, R., Allg
Sub 20 - Inaugural Deep Ea
/files/2020/02/Michelle_Sal

AuScope Distributed Acoustic Sensing (DAS) Collection

Distributed Acoustic Sensing (DAS) transforms fibre optic cable as one continuous sensing element with thousands of sensors at meter-spacing along the cable, measuring at a broad range of frequency (0.001 – 1000 Hz). A DAS interrogator sends laser signal pulses along the cable and continuously monitors the
phase and amplitude of the backscattered ligh
slightly stretched or compressed, and the dyna
phase of the backscattering. DAS arrays are ro
boreholes, urban telecommunication networks
new opportunities to revolutionize geophysical
source detection.

This data collection contains the time series, m
collected using AuScope-funded interrogators.

National ASTER Map of Australia

This is the parent datafile of a dataset that comprises a set of 14+ geoscience products made up of mosaiced ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) scenes across Australia. The individual geoscience products are a combination of bands and band ratios to highlight different mineral groups and parameters including: False colour composite CSIRO Landsat TM Regolith Ratios Green vegetation content Ferric oxide content Ferric oxide composition Ferrous iron index Opaque index AlOH group content AlOH group composition Kaolin group index FeOH group content MgOH group
content MgOH
Surface miner

--
This record re
Australia. See

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.

https://geonetwork.nci.org.au

Expanding our geophysical data collections on HPC

Geoscience Australia Geophysics Reference Data Collection

This collection has been compiled by Geoscience Australia from an extensive archive of over 2,200
geophysical surveys dating back to 1947. The datasets have been acquired by Geoscience Australia and
its State and Territory Governm
(magnetics and gravity) and ra
elevation datasets (digital elev
geophysical surveys, are also
on the individual surveys for e
radiometric and elevation surv
the national scale compilation
and are stored in both their fu
interpolated regular grid (raste
disciplines including geologica
environmental resource mana

AusPass Passive Seismic Collection

The Passive Seismic Data contains mirrored datasets hosted on the Australian Passive Seismic Server
(AusPass). AusPass is an initiative supported by funding from AuScope and the Australian National
University to host passive seismic (i.e. continuous waveform) data via the standardised FDSN Web Service
protocol. Data largely consi
individual stations, but also
network and mirrors compli
Seismic Network (IU), amor
in individual network metad
AusPass website and data

References:
Albuquerque Seismological
[Data set]. International Fed
Australian Research Data C
geophysics reference collec
Balfour, N.J., Salmon, M. &
Education, outreach, resear
https://doi.org/10.1785/022
Geoscience Australia. (202
September 2018). Commor
Salmon, M., Pickle, R., Allg
Sub 20 - Inaugural Deep Ea
/files/2020/02/Michelle_Sal

AuScope Distributed Acoustic Sensing (DAS) Collection

Distributed Acoustic Sensing (DAS) transforms fibre optic cable as one continuous sensing element with
thousands of sensors at meter-spacing along the cable, measuring at a broad range of frequency (0.001 –
1000 Hz). A DAS interrogator sends laser signal pulses along the cable and continuously monitors the
phase and amplitude of the backscattered ligh
slightly stretched or compressed, and the dyna
phase of the backscattering. DAS arrays are ro
boreholes, urban telecommunication networks
new opportunities to revolutionize geophysical
source detection.

This data collection contains the time series, m
collected using AuScope-funded interrogators.

National ASTER Map of Australia

This is the parent datafile of a dataset that comprises a set of 14+ geoscience products made up of
mosaiced ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) scenes across
Australia. The individual geoscience products are a combination of bands and band ratios to highlight
different mineral groups and parameters including: False colour composite CSIRO Landsat TM Regolith
Ratios Green vegetation content Ferric oxide content Ferric oxide composition Ferrous iron index Opaque
index AlOH group content AlOH group composition Kaolin group index FeOH group content MgOH group
content MgOH
Surface miner

--
This record re
Australia. See

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.

Please see Jo Croucher's talk on *"Vertical Integration of National Geophysical Data Assets to Support Next-generation Reproducible Research at Exascale"* @ 1400

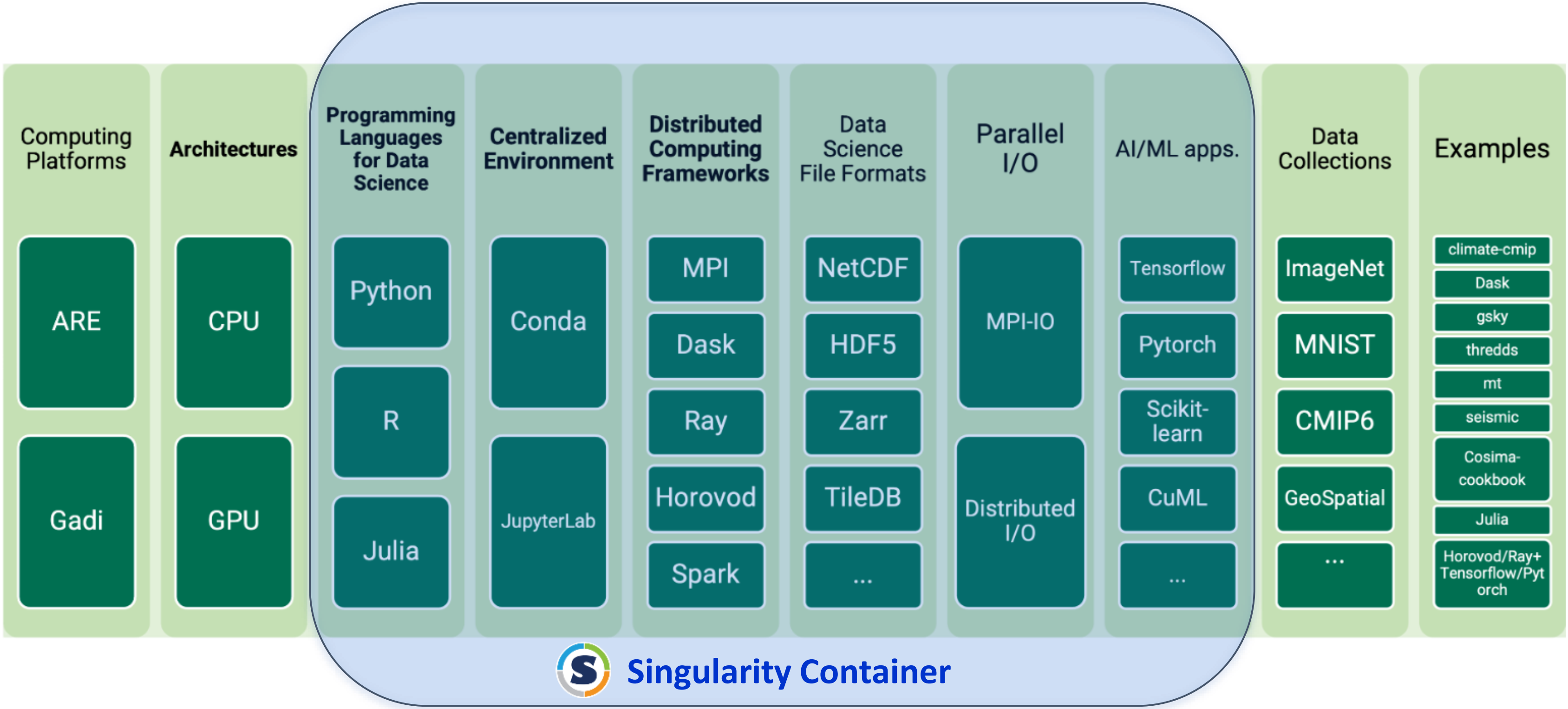
https://geonetwork.nci.org.au

Building up our geophysical software libraries on HPC

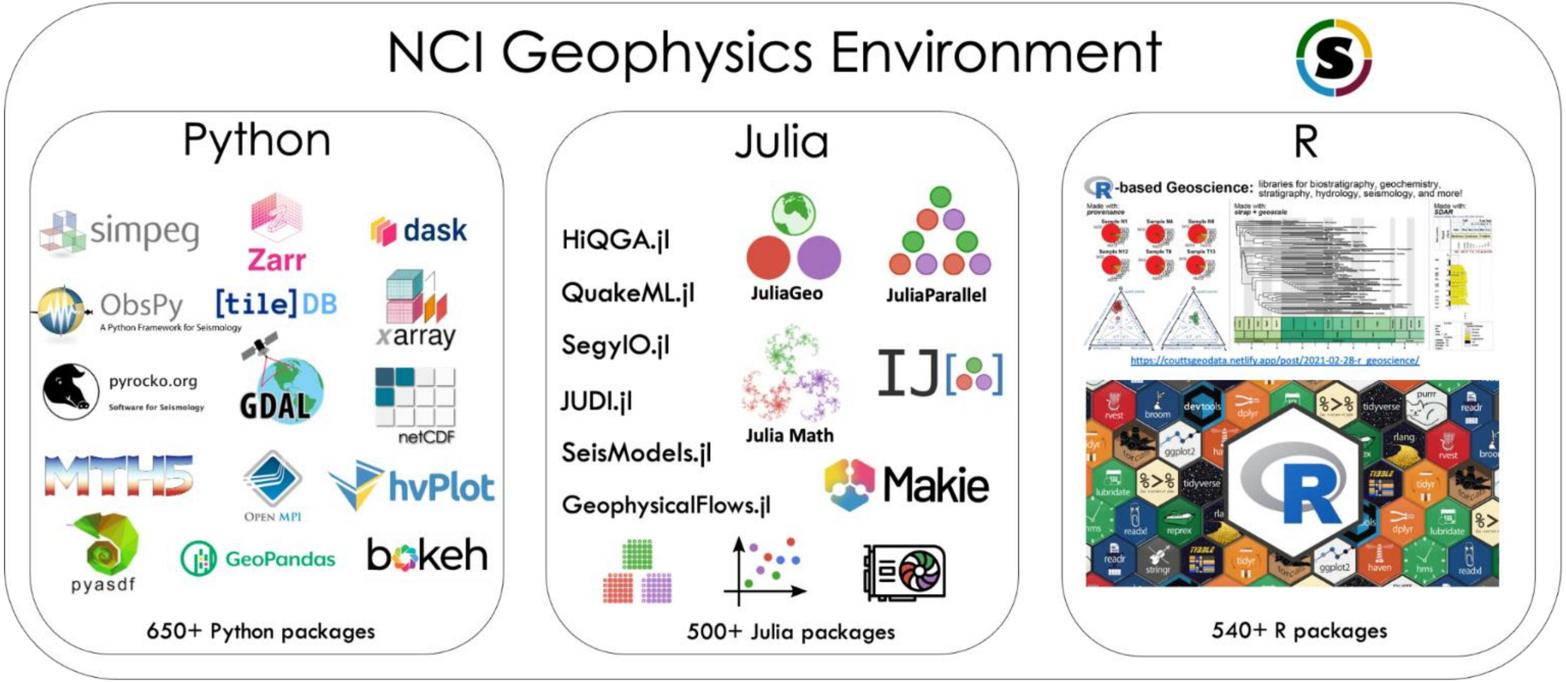
The following geophysical software are available on the NCI platforms mostly under NCI project up99:

- **BIRRP**: The Bounded Influence Remote Reference Processing program
- **EMTF**: Oregon State University robust single station, remote reference and multiple station MT timeseries data processing program
- **OCCAM1DCSEM / DIPOLE1D**: Smooth one-dimensional models from CSEM and MT data
- **OCCAM2DMT**: Occam's inversion for 2D MT modeling
- **MARE2DEM**: parallel adaptive finite element code for 2D forward and inverse modeling for electromagnetic geophysics
- **ModEM**: a flexible electromagnetic modelling and inversion program for 2D and 3D MT problems
  - https://my.nci.org.au/mancini/project/ModEM-geophys
- **FEMTIC**: A 3-D Finite Element MagnetoTelluric Inversion code
- **esys-escript**: a programming environment for implementing mathematical models in python using the finite element method (FEM).
- **Firedrake**: an automated system for solving partial differential equations using the finite element method (FEM).
- **jif3D**: a framework for joint inversion of different types of geophysical data in 3D.
- **OpenQuake**: an open-source application that allows users to compute seismic hazard and seismic risk of earthquakes on a global scale.

Specialised Software Environments: Development scope



Developing community driven specialised software environments



$ module use /g/data/up99/modulefiles
$ module load NCI-geophys/<version>

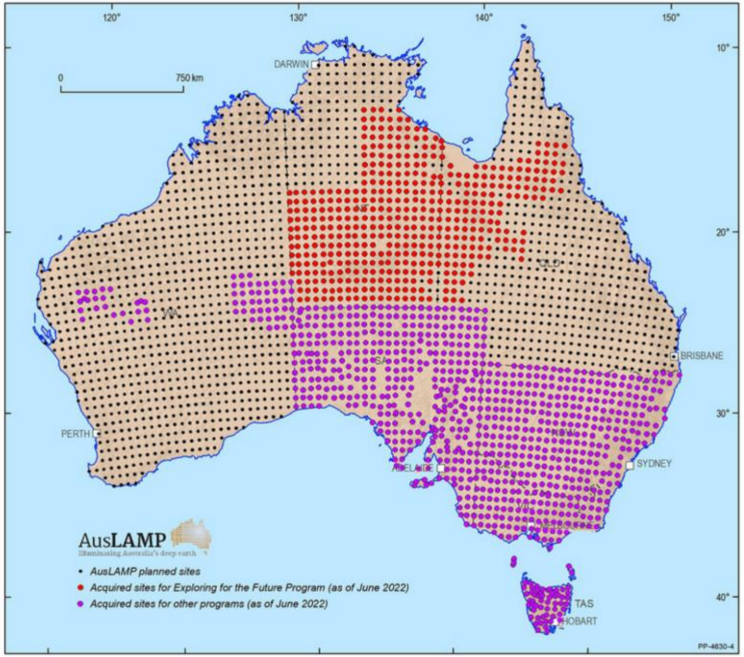
https://doi.org/10.6084/m9.figshare.21919584.v1

Machine Learning and data analysis software environments

	NCI-ai-ml	RAPIDS	NCI-data-analysis (including NCI dataset index)
Languages	Python	Python	Python, R, Julia
Packages	MPI, Dask, Ray, Horovod, Tensorflow, Pytorch, Keras, …	MPI, Dask, Ray, cuPy, cuDF, cuML, cuGraph, …	MPI, Dask, Ray, Spark, Xarray, NetCDF, Zarr, TileDB, intake, Scikit-learn, …
Processor Architecture	Deep learning on GPU	Classical learning on GPU	Machine learning on CPU

https://opus.nci.org.au/x/zAB0Bw

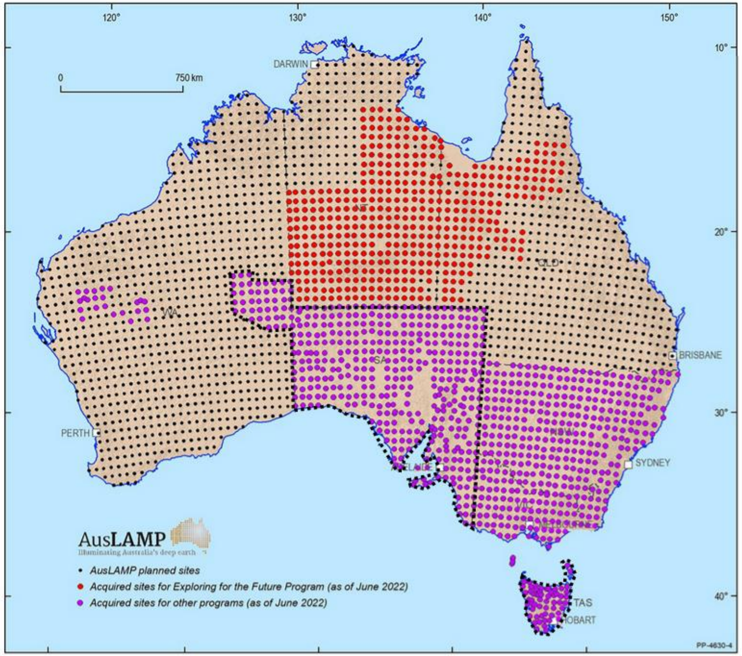
National Scale Analysis: The Australian Lithospheric Architecture Magnetotelluric Project (AusLAMP)



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

- A collaborative project between Geoscience Australia, the state and Northern Territory geological surveys, AuScope, universities and other research organisations.
- Aims to acquire long-period MT data at approximately 3000 sites across Australia.
- Dataset will be used to map electrical conductivity structure of the crust and upper mantle to improve the understanding of the geology and tectonic evolution of the Australian plate.
- Results may provide new information for identifying regions with mineral and energy resource potential.

National Scale Analysis: The Australian Lithospheric Architecture Magnetotelluric Project (AusLAMP)

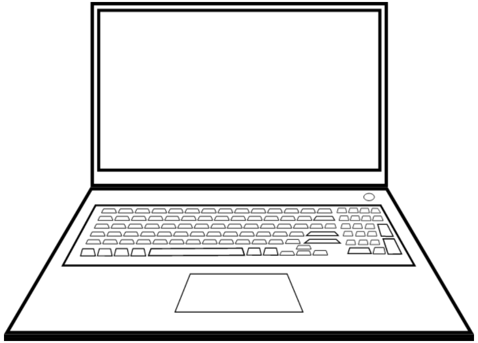


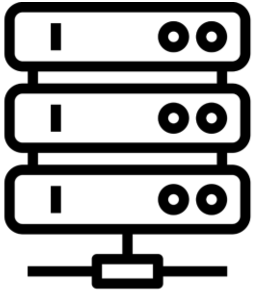
https://www.ga.gov.au/about/projects/resources/auslamp

- A collaborative project between Geoscience Australia, the state and Northern Territory geological surveys, AuScope, universities and other research organisations.
- Aims to acquire long-period MT data at approximately 3000 sites across Australia.
- Dataset will be used to map electrical conductivity structure of the crust and upper mantle to improve the understanding of the geology and tectonic evolution of the Australian plate.
- Results may provide new information for identifying regions with mineral and energy resource potential.

Can we process AusLAMP time series data at scale on HPC?

Where is the AusLAMP MT time series data?





Local Server

Can we process AusLAMP time series data at scale on HPC?

Where is the AusLAMP MT time series data?

PR6-24 Earth Data Logger

https://www.gfz-potsdam.de/en/section/geophysical-imaging/infrastructure/geophysical-instrument-pool-potsdam-gipp/pool-components/depas-pool/recorder-/-earthdata-pr6-24

LEMI 424

https://www.isr.lviv.ua/lemi424.htm

University of Adelaide Orange Boxes

Can we process AusLAMP time series data at scale on HPC?

Where is the AusLAMP MT time series data?

PR6-24 Earth Data Logger

```
3813265
3813266
3813275
3813281
3813275
3813277
3813277
3813272
3813277
3813281
3813278
3813281
3813287
3813284
3813282
3813282
3813284
3813290
3813290
3813288
3813290
```

LEMI 424

```
2016 10 27 00 00 00 27747.359 -1985.590 -47045.008  28.59  31.76  164.456  -45.318  -2.309  -2.227 14.24  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 01 27747.361 -1985.599 -47045.035  28.62  31.76  164.470  -45.330  -2.373  -2.357 14.23  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 02 27747.330 -1985.631 -47045.031  28.61  31.76  164.496  -45.316  -2.369  -2.305 14.24  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 03 27747.309 -1985.651 -47045.047  28.62  31.76  164.498  -45.307  -2.361  -2.297 14.24  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 04 27747.299 -1985.660 -47045.055  28.62  31.76  164.494  -45.273  -2.191  -2.182 14.25  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 05 27747.285 -1985.667 -47045.035  28.63  31.76  164.466  -45.232  -2.389  -2.215 14.25  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 06 27747.289 -1985.691 -47045.055  28.61  31.76  164.442  -45.215  -2.225  -2.160 14.21  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 07 27747.277 -1985.701 -47045.063  28.59  31.76  164.408  -45.168  -2.250  -2.275 14.25  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 08 27747.289 -1985.692 -47045.063  28.61  31.76  164.366  -45.112  -2.285  -2.229 14.25  485.8 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 09 27747.268 -1985.678 -47045.051  28.62  31.76  164.276  -45.098  -2.359  -2.213 14.25  485.8 2629.7502 S 12826.0778 E 9 2 2
2016 10 27 00 00 10 27747.293 -1985.671 -47045.066  28.59  31.76  164.181  -45.053  -2.285  -2.318 14.25  485.8 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 11 27747.285 -1985.654 -47045.059  28.59  31.76  164.079  -45.009  -2.443  -2.121 14.25  485.8 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 12 27747.303 -1985.644 -47045.059  28.60  31.76  164.041  -44.953  -2.338  -2.184 14.25  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 13 27747.307 -1985.628 -47045.055  28.62  31.76  163.971  -44.906  -2.342  -2.229 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 14 27747.320 -1985.614 -47045.055  28.64  31.76  163.931  -44.864  -2.385  -2.191 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 15 27747.336 -1985.617 -47045.039  28.63  31.76  163.912  -44.813  -2.408  -2.119 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 16 27747.340 -1985.622 -47045.023  28.60  31.76  163.919  -44.764  -2.373  -2.227 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 17 27747.363 -1985.599 -47045.023  28.61  31.76  163.886  -44.690  -2.348  -2.242 14.15  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 18 27747.400 -1985.600 -47045.027  28.62  31.76  163.901  -44.647  -2.318  -2.225 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 19 27747.412 -1985.590 -47045.008  28.59  31.76  163.924  -44.619  -2.369  -2.268 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 20 27747.426 -1985.571 -47045.000  28.61  31.76  163.910  -44.611  -2.391  -2.193 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 21 27747.449 -1985.591 -47044.980  28.63  31.76  163.923  -44.597  -2.219  -2.146 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 22 27747.469 -1985.583 -47044.973  28.61  31.76  163.931  -44.560  -2.307  -2.369 14.09  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 23 27747.473 -1985.569 -47044.980  28.64  31.76  163.908  -44.550  -2.336  -2.254 14.09  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 24 27747.506 -1985.585 -47044.977  28.64  31.76  163.923  -44.510  -2.307  -2.205 14.09  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 25 27747.516 -1985.583 -47044.965  28.62  31.76  163.926  -44.525  -2.322  -2.184 14.09  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 26 27747.510 -1985.579 -47044.957  28.59  31.76  163.919  -44.593  -2.396  -2.160 14.14  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 27 27747.523 -1985.588 -47044.973  28.63  31.76  163.888  -44.603  -2.188  -2.201 14.09  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 28 27747.525 -1985.580 -47044.957  28.62  31.76  163.860  -44.650  -2.266  -2.303 14.17  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 29 27747.527 -1985.580 -47044.969  28.63  31.76  163.838  -44.666  -2.344  -2.326 14.09  485.8 2629.7503 S 12826.0778 E 10 2 2
```

Can we process AusLAMP time series data at scale on HPC?

Where is the AusLAMP MT time series data?

PR6-24 Earth Data Logger

```
3813265
3813266
3813275
3813281
3813275
3813277
3813277
3813272
3813277
3813281
3813278
3813281
3813287
3813284
3813282
3813282
3813284
3813290
3813290
3813288
3813290
```

LEMI 424

```
2016 10 27 00 00 00 27747.359 -1985.590 -47045.008  28.59  31.76  164.456  -45.318  -2.309  -2.227 14.24  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 01 27747.361 -1985.599 -47045.035  28.62  31.76  164.470  -45.330  -2.373  -2.357 14.23  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 02 27747.330 -1985.631 -47045.031  28.61  31.76  164.496  -45.316  -2.369  -2.305 14.24  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 03 27747.309 -1985.651 -47045.047  28.62  31.76  164.498  -45.307  -2.361  -2.297 14.24  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 04 27747.299 -1985.660 -47045.055  28.62  31.76  164.494  -45.273  -2.191  -2.182 14.25  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 05 27747.285 -1985.667 -47045.035  28.63  31.76  164.466  -45.232  -2.389  -2.215 14.25  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 06 27747.289 -1985.691 -47045.055  28.61  31.76  164.442  -45.215  -2.225  -2.160 14.21  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 07 27747.277 -1985.701 -47045.063  28.59  31.76  164.408  -45.168  -2.250  -2.275 14.25  485.9 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 08 27747.289 -1985.692 -47045.063  28.61  31.76  164.366  -45.112  -2.285  -2.229 14.25  485.8 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 09 27747.268 -1985.678 -47045.051  28.62  31.76  164.276  -45.098  -2.359  -2.213 14.25  485.8 2629.7502 S 12826.0778 E 9 2 2
2016 10 27 00 00 10 27747.293 -1985.671 -47045.066  28.59  31.76  164.181  -45.053  -2.285  -2.318 14.25  485.8 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 11 27747.285 -1985.654 -47045.059  28.59  31.76  164.079  -45.009  -2.443  -2.121 14.25  485.8 2629.7502 S 12826.0778 E 10 2 2
2016 10 27 00 00 12 27747.303 -1985.644 -47045.059  28.60  31.76  164.041  -44.953  -2.338  -2.184 14.25  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 13 27747.307 -1985.628 -47045.055  28.62  31.76  163.971  -44.906  -2.342  -2.229 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 14 27747.320 -1985.614 -47045.055  28.64  31.76  163.931  -44.864  -2.385  -2.191 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 15 27747.336 -1985.617 -47045.039  28.63  31.76  163.912  -44.813  -2.408  -2.119 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 16 27747.340 -1985.622 -47045.023  28.60  31.76  163.919  -44.764  -2.373  -2.227 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 17 27747.363 -1985.599 -47045.023  28.61  31.76  163.886  -44.690  -2.348  -2.242 14.15  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 18 27747.400 -1985.600 -47045.027  28.62  31.76  163.901  -44.647  -2.318  -2.225 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 19 27747.412 -1985.590 -47045.008  28.59  31.76  163.924  -44.619  -2.369  -2.268 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 20 27747.426 -1985.571 -47045.000  28.61  31.76  163.910  -44.611  -2.391  -2.193 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 21 27747.449 -1985.591 -47044.980  28.63  31.76  163.923  -44.597  -2.219  -2.146 14.10  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 22 27747.469 -1985.583 -47044.973  28.61  31.76  163.931  -44.560  -2.307  -2.369 14.09  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 23 27747.473 -1985.569 -47044.980  28.64  31.76  163.908  -44.550  -2.336  -2.254 14.09  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 24 27747.506 -1985.585 -47044.977  28.64  31.76  163.923  -44.510  -2.307  -2.205 14.09  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 25 27747.516 -1985.583 -47044.965  28.62  31.76  163.926  -44.525  -2.322  -2.184 14.09  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 26 27747.510 -1985.579 -47044.957  28.59  31.76  163.919  -44.593  -2.396  -2.160 14.14  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 27 27747.523 -1985.588 -47044.973  28.63  31.76  163.888  -44.603  -2.188  -2.201 14.09  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 28 27747.525 -1985.580 -47044.957  28.62  31.76  163.860  -44.650  -2.266  -2.303 14.17  485.8 2629.7503 S 12826.0778 E 10 2 2
2016 10 27 00 00 29 27747.527 -1985.580 -47044.969  28.63  31.76  163.838  -44.666  -2.344  -2.326 14.09  485.8 2629.7503 S 12826.0778 E 10 2 2
```

NCI AUSTRALIA

Can we process AusLAMP time series data at scale on HPC?

Where is the AusLAMP MT time series metadata?

1	Not Used	Not Used	Not Used	Not Used	Compulsory	Optional	Optional	Optional	Compulsory	Compulsory	Compulsory	Compulsory	Compulsory	Compulsory	Compulsory	Optional
2	Site Number	Coords Available	Deployed Date Entered	Pickedup Date Entered	Site Number	UTC Start Time Minute	UTC Start Time Second	Deployment Julian Day	Recording Method	MT Recorder Type/ Model	Magnetometer Type/ Model	Electrode Type/ Model	Power Source Type/ Model	Data Confidentiality	North Reference	Drift calculation end time - Time
406	SA333	Coords	dep date	Retrival date	SA333	44	12	139	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
407	SA333-2A	Coords	dep date	Retrival date	SA333-2A	23	0	170	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
408	SA333-2B	Coords	dep date	Retrival date	SA333-2B	56	0	170	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
409	SA334	Coords	dep date	Retrival date	SA334	55	0	139	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
410	SA335	Coords	dep date	Retrival date	SA335	53	51	140	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
411	SA336	Coords	dep date	Retrival date	SA336	17	24	140	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
412	SA337	Coords	dep date	Retrival date	SA337	6	48	142	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
413	SA338	Coords	dep date	Retrival date	SA338	3	17	142	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
414	SA339	Coords	dep date	Retrival date	SA339	26	51	143	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
415	SA340	Coords	dep date	Retrival date	SA340	35	0	140	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
416	SA340-2	Coords	dep date	Retrival date	SA340-2	1	0	143	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
417	SA341	Coords	dep date	Retrival date	SA341	10	0	324	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
418	SA342	Coords	dep date	Retrival date	SA342	54	0	324	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
419	SA344	Coords	dep date	Retrival date	SA344	35	0	135	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
420	SA344-2	Coords	dep date	Retrival date	SA344-2	6	16	171	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
421	SA345	Coords	dep date	Retrival date	SA345	59	15	132	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
422	SA346	Coords	dep date	Retrival date	SA346	39	0	132	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
423	SA347	Coords	dep date	Retrival date	SA347	7	0	234	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
424	SA348	Coords	dep date	Retrival date	SA348	21	0	232	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
425	SA349	Coords	dep date	Retrival date	SA349			231	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
426	SA350	Coords	dep date	Retrival date	SA350	41	13	231	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
427	SA351	Coords	dep date	Retrival date	SA351	52	20	231	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	
428	SA354	Coords	dep date	Retrival date	SA354	18	52	79	LP	Earth Data Recc	Bartington, Mo	SDEC (France), N	12Volt 72 Amp/Hr Battery, Power Supply Charging - Solar Panel, 60Watt	Open	Magnetic North	

Converting MT time series (meta)data to international standards

ELSEVIER

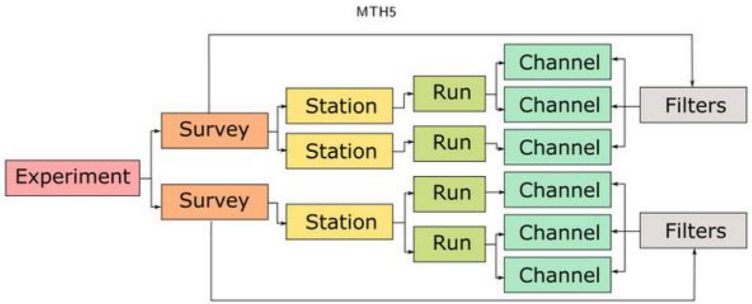
Computers & Geosciences

Volume 162, May 2022, 105102

Research paper

MTH5: An archive and exchangeable data format for magnetotelluric time series data

Jared Peacock [a], Karl Kappler [b], Lindsey Heagy [c], Timothy Ronan [d], Anna Kelbert [e], Andrew Frassetto [d]



MT_metadata

- Current version of mt_metadata has 355 attributes
  - 152 required parameters
  - 203 non mandatory parameters (but useful to have)

- For legacy AusLAMP metadata, we could not meet this standard as many of the required parameters were not recorded at the time of acquisition

- We did our best with what we had and worked on a Legacy AusLAMP MT_metadata profile that was a 72 element subset of the full MT_metadata standard

- Held regular meetings with the USGS based lead developer of MT_metadata and added in some revisions to the original standard (e.g., survey.funding_source.organization, survey.state, more options for station.release_license)

Scaling up MT time series processing

The **M**agneto**T**ellurics **t**ime **s**eries **d**ata **p**ublication (**MTtsdp**) codes: https://github.com/nci/MTtsdp

Processing Levels	Name	Description
Packed Raw Data	Raw Time Series	Telemetry data streamed from site loggers
Level 0	Edited Time Series	Time ordered instrument recorded data (e.g., raw voltages, counts) at full resolution
Level 1	Transformed Time Series	Level 0 data that have been transformed (e.g., calibrated, resampled, rotated, had noisy data removed, filters applied).
Level 2	Derived frequency domain processed data	Geophysical parameters (e.g., impedance tensors) derived from frequency domain time series processing of Level 1 data
Level 3	Derived modelling inputs and outputs	Level 2 parameters converted into input files for modelling and inversion algorithms with outputs mapped onto space-time grids.

Rees, N., Evans, B., Heinson, G., Conway, D., Yang, R., Thiel, S., Robertson, K., Druken, K., Goleby, B., Wang, J., Wyborn, L. & Seillé, H., 2019. The Geosciences DeVL Experiment: new information generated from old magnetotelluric data of The University of Adelaide on the NCI High Performance Computing Platform, ASEG Extended Abstracts, 2019:1, 1-6, DOI: 10.1080/22020586.2019.12073015

Scaling up MT time series processing

The **M**agneto**T**ellurics **t**ime **s**eries **d**ata **p**ublication (**MTtsdp**) codes: https://github.com/nci/MTtsdp

Processing Levels	Name	Description
Packed Raw Data	Raw Time Series	Telemetry data streamed from site loggers
Level 0	Edited Time Series	Time ordered instrument recorded data (e.g., raw voltages, counts) at full resolution
Level 1	Transformed Time Series	Level 0 data that have been transformed (e.g., calibrated, resampled, rotated, had noisy data removed, filters applied).
Level 2	Derived frequency domain processed data	Geophysical parameters (e.g., impedance tensors) derived from frequency domain time series processing of Level 1 data
Level 3	Derived modelling inputs and outputs	Level 2 parameters converted into input files for modelling and inversion algorithms with outputs mapped onto space-time grids.

Rees, N., Evans, B., Heinson, G., Conway, D., Yang, R., Thiel, S., Robertson, K., Druken, K., Goleby, B., Wang, J., Wyborn, L. & Seillé, H., 2019. The Geosciences DeVL Experiment: new information generated from old magnetotelluric data of The University of Adelaide on the NCI High Performance Computing Platform, ASEG Extended Abstracts, 2019:1, 1-6, DOI: 10.1080/22020586.2019.12073015

HPC has reduced processing times significantly

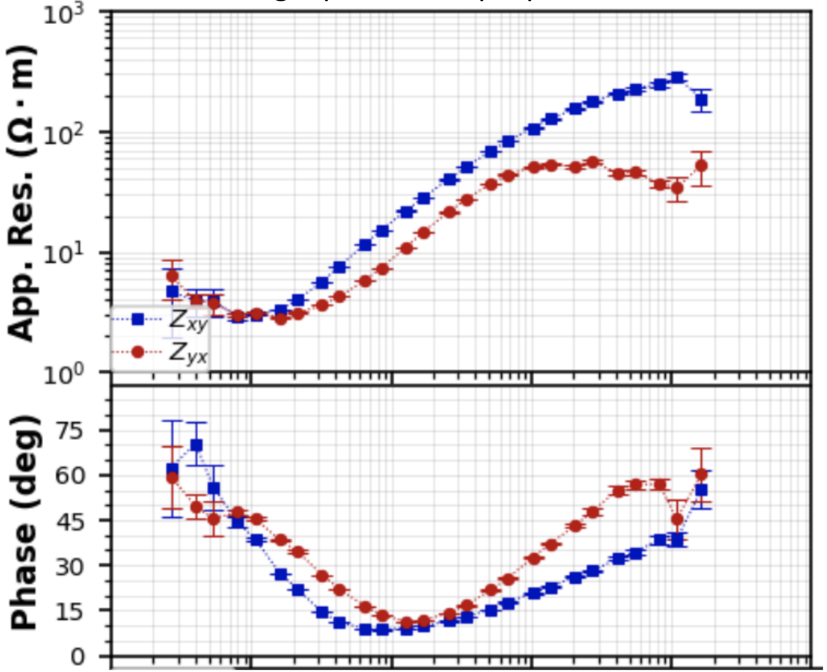
Parallel I/O via NCI-geophys environment

- MTH5: HDF5 data container for magnetotelluric time series data
- MT_metadata: standardised time series and transfer function metadata for magnetotelluric data
- Tests conducted on the AusLAMP Musgraves Province time series data: https://dx.doi.org/10.25914/58gr-1550

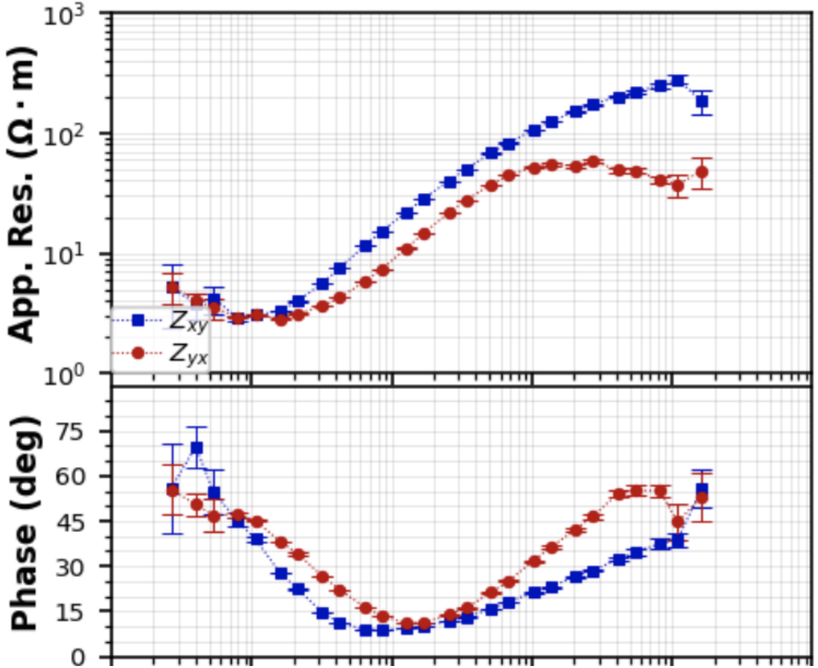
Dataset of 93 stations	Serial I/O	MPI based Parallel I/O (96 cores)
Level 0: one mth5/mt_metadata file per station	~ 5 hours 47 minutes	~ 4 minutes
Level 1: one mth5/mt_metadata file per station	~ 49 minutes	~1 minute 13 seconds
Level 2: one EDI file per station	~ 2 hours 30 minutes	~ 2 minutes

Computational reproducibility

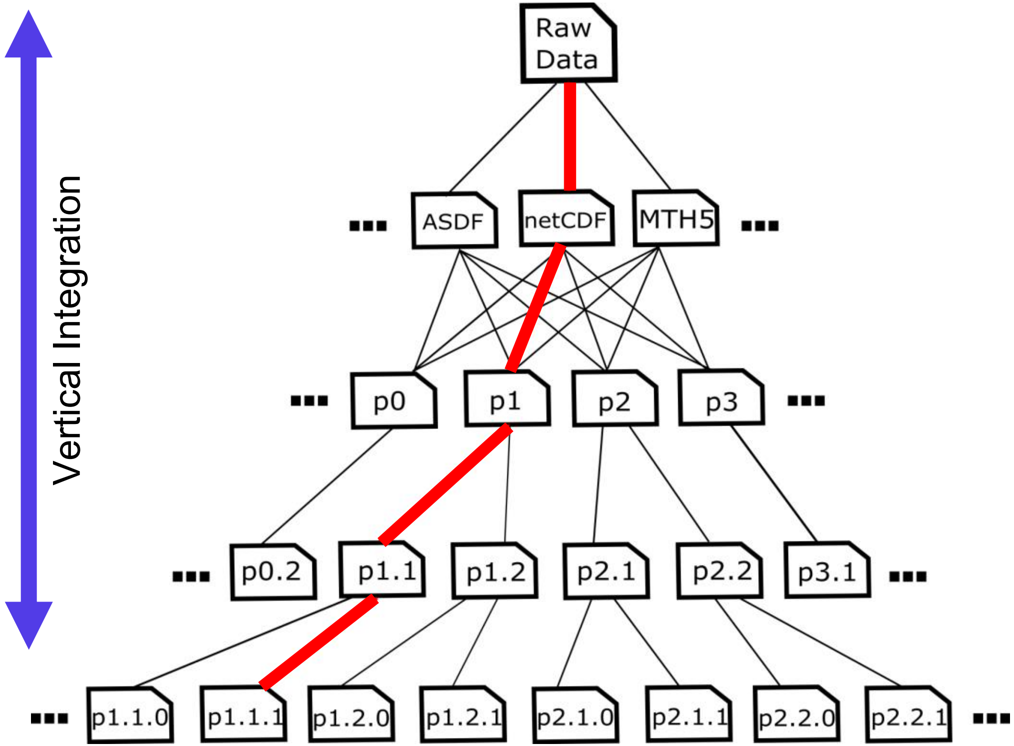
Original station WA72 processed using a personal laptop



Processing station WA72 using the MTtsdp automations on HPC



Transparent provenance between processing levels



Level 0: Edited raw time series magnetotelluric (MT) survey data in different formats

Connected via PIDs

Level 1: Calibrated, resampled MT time series data

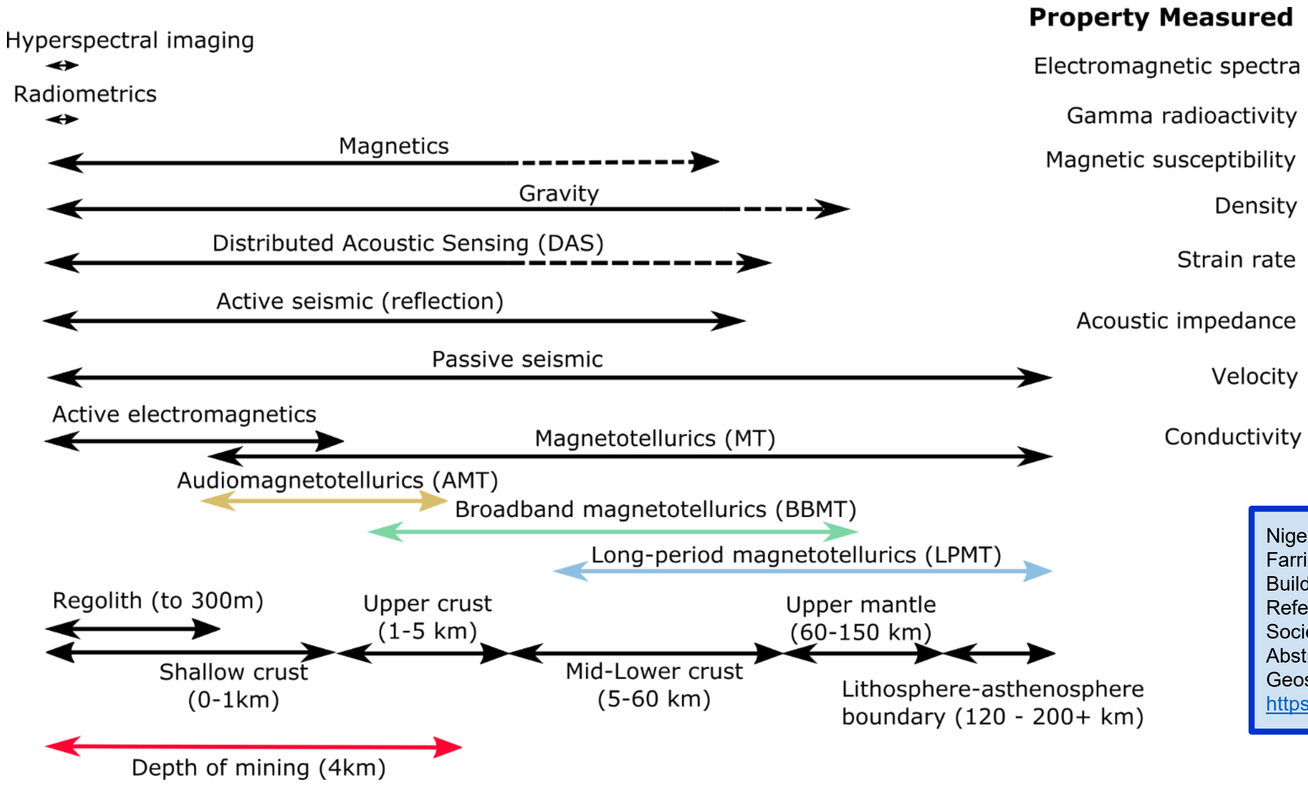
Connected via PIDs

Level 2: Derived frequency domain processed data

Connected via PIDs

Level 3: MT Modelling input/output files

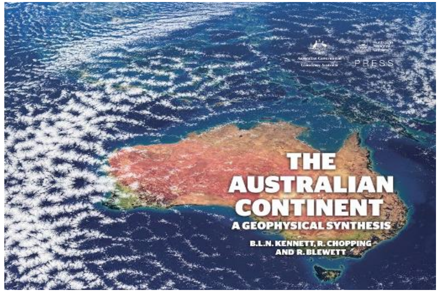
2030 ambition: Multiphysics Analysis but starting with PS and MT



Types of geophysical data collected in Australia, the physical property measured and the depth of the crust that is sampled: also shown is the depth of current mining. Figure modified from original of Richard Chopping (GSWA).

Nigel Rees, Lesley Wyborn, Ben Evans, Rebecca Farrington, Tim Rawling, Rui Yang, & Yue Sun. (2023). Building a National High-Resolution Geophysics Reference Collection for 2030 Computation. Australian Society of Exploration Geophysicists Extended Abstracts, Volume 2023, 4th Australasian Exploration Geoscience Conference, Brisbane, 2023. https://doi.org/10.5281/zenodo.7980192

Can we create a collaborative National High Resolution Geophysics Data Platform?



Brian Kennett, Richard Chopping and Richard Blewett, 2018. The Australian Continent, a Geophysical synthesis. Available on https://press.anu.edu.au/publications/australian-continent#tabanchor

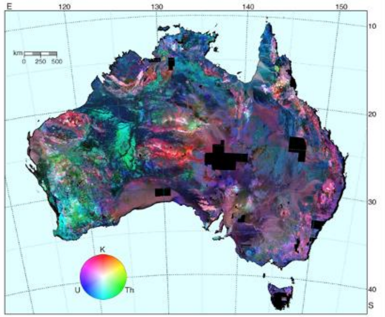


Figure 3.1: Gamma-ray spectrometric map for the continent of Australia created by

Radiometrics

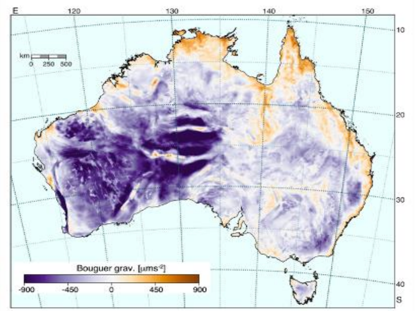


Figure 5.7: Bouguer gravity anomaly.

Bouger Gravity

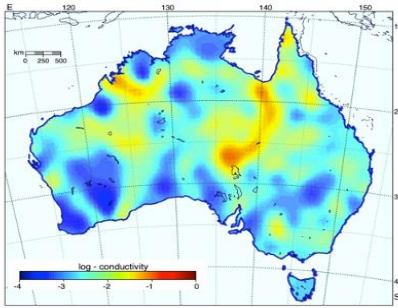


Figure 10.3: Logarithm of electrical conductivity [S/m] at a depth of 52 km.

Electrical Conductivity at 52 km

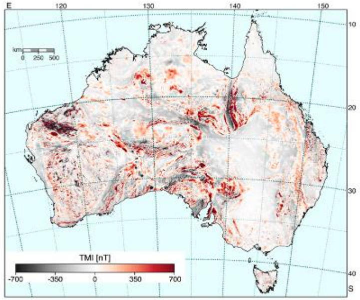


Figure 4.6: Total magnetic intensity across the Australian continent.

Total Magnetic Intensity

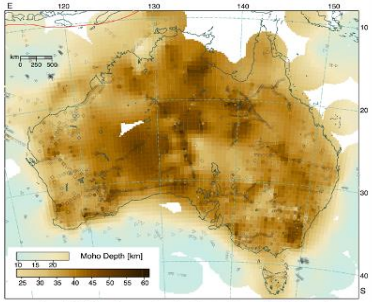


Figure 7.16: Moho surface across Australia utilising the full range of seismic information

Moho Depth

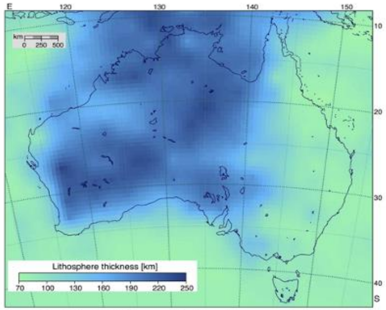


Figure 8.9: Estimate of thickness of the lithosphere across the Australian region.

Thickness of the lithosphere

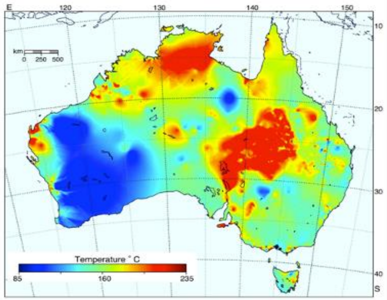


Figure 11.2: OZtemp estimate of the temperature distribution at 5 km depth across the Australian continent.

Temperature at 5km

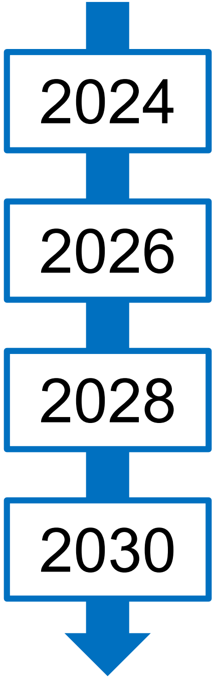
Preparing for 2030: we no longer have the gift of time

Photo by Nathan Dumlao on Unsplash

We need to make up our minds today if we want to scale up.

If so we need to:

- Find all the rawer forms of critical geophysical datasets (and their collectors!) and work on ensuring all metadata, data and vocabularies are FAIR and machine actionable.
- Start working now on more automated systems that capture provenance through each successive processing level.
- Make less processed forms of data more accessible and able to be aggregated into seamless national high-resolution datasets.
- Ensure that whatever we do, it is always scalable to the future and can maximise benefits from new compute, data and software technologies as they come on line.



NCI Contacts

General enquiries: +61 2 6125 9800

Support: help@nci.org.au

Email: nigel.rees@anu.edu.au

Address

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

License