Modernising Australian geophysics datasets to enable National-scale multiphysics computation at Exascale: one size can no longer fit all.

Lesley Wyborn[1,2]
lesley.wyborn@anu.edu.au

Nigel Rees[1]
nigel.rees@anu.edu.au

Rebecca Farrington[3]
rebecca@auscope.org.au

Ben Evans[1]
ben.evans@anu.edu.au

Hannes Hollmann[1]
hannes.hollmann@anu.edu.au

Jo Croucher[1]
jo.croucher@anu.edu.au

Rui Yang[1]
rui.yang@anu.edu.au

Tim Rawling[3]
Tim@auscope.org.au

[1]National Computational Infrastructure, Australian National University, Canberra, ACT.
[2]Research School of Earth Sciences, Australian National University, Canberra, ACT.
[3]AuScope, School of Earth Sciences, University of Melbourne, Victoria.

SUMMARY

Although available computational power has increased significantly in the last decade, the geophysics community has been slow to utilise the latest High Performance Computing (HPC) resources. The 2030 Geophysics Collections Project, a collaboration between the National Computational Infrastructure (NCI), AuScope, Australian Research Data Commons (ARDC) and Terrestrial Ecosystem Research Network (TERN), investigated what was required to make existing Australian geophysical data collections compliant with the FAIR Principles, and openly accessible on existing HPC resources at NCI, and ultimately scaling to Exascale resources by 2030. A selection of AuScope-funded time series datasets in Magnetotellurics (MT), Passive Seismic (PS) and Distributed Acoustic Sensing (DAS) were ingested and organised on NCI's Gadi Tier 1 supercomputer to facilitate multi-physics (re)processing, modelling and analysis at scale. For MT, optimisation for HPC required reformatting datasets into the new MTH5/ mt_metadata standard, with benchmark processing showing that generating different MT processing levels for entire surveys can now be done transparently and in a matter of minutes utilising current HPC capability. In parallel to these data activities, the NCI-geophysics and NCI-AI-ML software environments were developed to enable a seamless experience for geophysical data analysis, processing and modelling on HPC and to help lower the barrier to entry for geophysical researchers wanting to perform AI/ML analysis utilising Gadi GPU resources.

The 2030 Geophysics Collections Project showed that to realise the full value of significant investments going back over 50 years in acquiring continental-scale geophysical datasets in government, industry and academia, we need to consider a collaborative effort from all three sectors to remaster our current geophysical assets into modern HPC-compatible, self-describing formats suitable for in-situ processing to enable high-resolution processing at scale, and readiness for new computational techniques. However, there is still a need to make datasets accessible in current formats: and hence one size can no longer fit all.

Key words: HPC, self-describing formats, rich metadata, co-located data and compute.

INTRODUCTION

Although most people are aware of the exponentially growing volumes of data and the concept of the Data Deluge (Hey and Trefethen, 2003) or Data Tsunami, what is often not appreciated is that simultaneously the capacity of computational infrastructures available to process the vast amounts of geophysical data acquired in Australia by the government, industry and research sectors have also been steadily increasing. To illustrate this, in 1993 the largest computer was 124 GigaFlop/s (Top500 1993); in June 2024 it was 1.206 ExaFlop/s (Top500 2024). However, in many areas of geophysics, we have yet to fully utilise this increased computational capacity.

Many geophysical acquisition programs acquire datasets that are at least terabytes in volume for a single experiment or survey. Additionally, consolidating and remastering existing datasets into large scale collections can amount to Petabytes in volume. However, the dominant paradigm of finding and accessing geophysical datasets is to download files for local processing, which limits accessible files to tens of Gigabytes, resulting in datasets being subsampled, generalised or broken up into smaller files to create download friendly datasets for efficient delivery. Similarly, datasets are downsized to fit the capability of either the infrastructures to store and deliver them and/or the actual size of the datasets that processing software can analyse (Wyborn and Woodcock, 2010). Further, many geophysical disciplines are often using standards and processing algorithms developed over 30 years ago (e.g., the Electrical Data Interchange (EDI) standard for magnetotelluric data developed by Wight (1987)): many of these standards are not optimised for modern data processing.

In comparison to climate, astronomy and genomics, the geosciences have been slow to evolve to accessing compute and data resources on the cloud and/or undertaking in situ processing on dedicated High Performance Computing (HPC) facilities, particularly for really large volume datasets. Possible reasons for this slow uptake are 1) traditional ASCII-based geophysics standards with 'side-car' metadata files (Ip et al., 2019) are not easy to use in HPC environments; 2) many data provider's infrastructures (storage, bandwidth, etc.) make it difficult to access and share large volume, high resolution time series data; and 3) many publicly available software and analytic tools to process geophysics data only read the old standards and/or are not capable of processing large volume data and/or high resolution files. Even with access to public clouds, the data-intensive nature of geophysical processing methods results in prohibitively high egress charges, and in some areas, the carbon footprint of processing times is becoming an increasing concern.

The recently completed 2030 Geophysics Collections Project (Rees et al., 2023; Evans et al., 2024), a collaboration between the National Computational Infrastructure (NCI), AuScope, Australian Research Data Commons (ARDC) and Terrestrial Ecosystem Research Network (TERN) set out to investigate what were the infrastructures required to support next-generation geophysical research at scale. The focus firstly was on ensuring existing raw and high-resolution versions of datasets and software would be compliant with the Findable, Accessible, Interoperable and Reusable (FAIR, Wilkinson et al., 2016) principles and would be suitable for programmatic access on existing HPC infrastructures at Petascale and secondly, to investigate whether these methods would further scale to Exascale environments.

The data trialled in the 2030 Geophysics Collections Project were the AuScope funded South Australian, Musgraves and Tasmanian Magnetotelluric (MT) datasets of the Australian Lithospheric Architecture Magnetotelluric Project (AusLAMP) project (The University of Adelaide, 2020), the AusPASS Passive Seismic (PS) collection (Salmon et al., 2020) and Distributed Acoustic Sensing (DAS) datasets from the Research School of Earth Sciences at the Australian National University (AuScope, 2023). Much of this data was acquired with old instruments and was collected prior to modern data infrastructures, standards, and formats. For MT, in many cases these datasets in their raw and minimally processed formats (L0, L1) are stored offline on tapes, CDs and hard drives and only highly evolved data products (L2, L3) are accessible online. Where the time series datasets are available, it is usually a requirement to contact the data provider directly to be able to obtain the data. The inability to easily access and reuse the time series data stifles innovation as users of the datasets are forced to rely on the time series processing of the data provider, which may or may not suit the desired use cases and inhibits users from developing new processing algorithms that are fine tuned to their local parameters.

This paper will present the key findings from this work based mainly on the Musgraves MT datasets of the AusLAMP Project (The University of Adelaide, 2022) and will note parallels in Passive Seismic datasets from the AusARRAY project. The paper will then offer solutions to modernising Australian Geophysics datasets to increase uptake in HPC environments, lay the foundation for multiphysics national-scale inversions and modelling at higher resolution, and increase the uptake and sophistication of Artificial Intelligence (AI) and Machine Learning (ML) techniques.
.

THE 2030 GEOPHYSICS COLLECTIONS PROJECT: TESTING THE POTENTIAL OF PROCESSING AT EXASCALE

In the 2030 Geophysics Collections Project, a selection of AuScope funded minimally processed MT, PS and DAS time series datasets were ingested and organised on NCI's Gadi Tier 1 research supercomputer to facilitate multi-physics (re)processing, modelling and analysis at scale with the computational tools available within the NCI. Further details are available in Rees et al. (2023) and Evans et al. (2024).

One of the major challenges encountered in the data work packages was accessing the original data collected in the field. Limited connectivity at remote field sites in Australia often led to the collection of field data on hard drives. Unfortunately, storage constraints at some agencies and institutions resulted in a significant portion of the larger volume time-series data being persistently stored on offline tapes, CDs and hard drives. An Australia-wide "scavenger hunt" was required to locate all the original field data and associated metadata for some MT survey data. Anecdotal evidence suggests that many raw and minimally processed geophysical assets are stored this way. An additional concern is that much of the field-based metadata is recorded in hard copy field notebooks, and in some cases, this metadata is stored in EXCEL files separate from the machine-observed data. Modern geophysical metadata standards have evolved to be much richer and more explicit, as noted by Kelbert (2020). If compliant metadata was not initially provided with the observed data, efforts were made to retrieve the original field log sheets and consult with the original collectors of the data to help fill in any gaps in the metadata (see also Kelbert et al., 2018).

Once all the raw MT field data and metadata were located, a committed effort went into building automation that 1) generated different MT processing levels at scale and 2) converted existing raw MT instrument time series data into the new international high-performant and self-describing MTH5/mt_metadata standard (Peacock et al., 2022). Benchmark experiments show that generating different MT processing levels for entire surveys can now be done transparently and in a matter of minutes on HPC (Rees et al., 2023).

However, modernising the data alone is not sufficient and in parallel, the NCI-geophysics and NCI-AI-ML software environments were developed to streamline the analysis, processing, and modelling of geophysical data on HPC. The NCI-geophysics module compiles many different complex, intricate and scalable open source geophysical and data science related libraries written in Python, Julia and R, whilst the NCI-AI-ML software module has bundled together popular machine learning packages (including Pytorch, Tensorflow, Horovod) and configured them for HPC, thus lowering the barrier to entry for geophysical researchers wanting to perform AI/ML analysis and workflows. These managed software environments allow users the ability to fluently scale their analysis to NCI's HPC Gadi and Australian Research Environment (ARE) systems utilising CPU and GPU resources. Standalone open-source geophysics related applications that support parallel computing have also progressively been added to Gadi.

IMPLICATIONS AND UPTAKE OF THE 2030 GEOPHYSICS COLLECTIONS PROJECT

The 2030 Geophysics Collections Project revealed the critical nature of transforming raw geophysical data into modern HDF5-based formats in order to enhance its useability on HPC systems. However, merely converting to new high performant data formats is not enough; these datasets must also be integrated with well managed and scalable software environments on HPC. These software ecosystems enable users to efficiently access and analyse the data, as well as develop their own customised parallel workflows that harness the full power of HPC. By providing a comprehensive infrastructure, geophysics researchers are empowered to efficiently leverage data in HPC settings and tailor multi-physics workflows to their specific needs, thereby accelerating the pace of research innovation. An added bonus of the integrated environment is that it supports greater scientific transparency and reproducibility.

While the MT community is making the transition to HDF5 based formats, other geophysical communities have yet to fully embrace HPC-enabled standards and formats. Although legacy standards may still support current work practices and applications, scaling to data-intensive methods will require the move to more modern self-describing formats. For example, the international passive seismic community is using several candidate modern self-describing standards, e.g., the Adaptable Seismic Data Format (ASDF, Kirsher et al., 2016; Gorbatov et al., 2020; Hassan et al., 2020) and TileDB (Bergland et al., 2022; Trabant et al., 2022). Despite this progress, many passive seismic communities continue to adhere to the miniSEED format, and hence for the foreseeable future, data will need to be available in both traditional and in modern self-describing formats.

Usage of DAS technology is increasing in geophysical applications and there is a need to implement DAS-specific standards to facilitate the integration of DAS measurements across experiments and increase reusability. The sheer volume of raw DAS data poses a significant computational challenge for the seismological community, both in terms of processing and disseminating this data. There is still uncertainty in best practices for describing, accessing and sharing DAS data. Recent efforts have seen the proposal of a new metadata standard by Hui Lai et al. (2024), along with the development of cloud-optimised data platforms utilising object storage and cloud-optimised data formats like Zarr and TileDB, as detailed by Ni et al. (2024).

There is also an initiative led by the United States Geological Survey (USGS) to convert airborne geophysical data surveys such as electromagnetic, magnetic and radiometric into the new Geophysical Survey (GS) data standard and an open-source Python toolbox (GSPy) (James et al., 2022). The GS convention leverages the well-established and widely used NetCDF file format and builds on the Climate and Forecasts (CF) metadata convention. Already, private geophysical acquisition and software companies have reached out to learn more about the GS standard as they are considering how they can incorporate the GSPy standard into their workflows.

Internationally, the major geophysical repositories and projects building Petascale and pre-Exascale software development projects include Geosphere INfrastructures for QUestions into Integrated REsearch (Geo-INQUIRE, Cotton et al., 2023), European Plate Observing System (EPOS, Cocco et al., 2022), Centre of Excellence in the domain of Solid Earth (ChEESE, Folch et al., 2023) and EarthScope (Zawacki et al., 2023) who are now migrating some of their datasets to the new HPC formats and/or new infrastructures.

The 2030 Geophysics Collections Project also highlighted the need to distinguish between working formats and delivery formats. Currently, many geophysical data producers deliver geophysical data in formats tailored for efficient delivery and streaming, often in the form of images, data products and at times in proprietary formats. However, formats suitable for online GIS systems and dashboards are not necessarily suitable for data-intensive processing on HPC systems.

To increase the uptake of geophysics data on HPC, there is a need to prioritise the standardisation and adoption of self-describing formats such as MTH5, ASDF, HDF5, NetCDF, Zarr, TileDB, etc. The endorsement of these standards by National Societies such as the Australian Society of Exploration of Geophysicists (ASEG) and various Associations within the International Union of Geodesy and Geophysics (IUGG) is seen as critical. The irony is that the HDF format was developed initially as a data model, file format and I/O library designed for storing, exchanging, managing, and archiving complex data including scientific, engineering, and remote sensing data (Folk et al., 1999). The HDF format is more efficient in terms of volume when compared to ASCII-based formats, allowing HDF-based formats to effectively serve a dual purpose - archiving geophysical datasets and enabling the sharing and managing of them for in situ access.

CONCLUSIONS AND FUTURE PLANS

Although the 2030 Geophysics Collections Project highlighted the difficulties in converting legacy geophysics data collected prior to 2020, it also showed the rewards in making rawer geophysical data FAIR compliant and available on HPC with managed scalable software environments that support greater scientific transparency and reproducibility. This infrastructure empowers researchers to efficiently develop their own HPC multi-physics workflows tailored to their specific use-cases, thereby accelerating the pace of research innovation.

The HPC user community at Petascale and Exasale is growing in the research sector, particularly in Europe where the emphasis is on using geophysics in hazard prediction for faster than real-time and higher resolution impact maps (e.g., Folch et al., 2023). The petroleum industry has also moved towards HPC more than a decade ago. It is now time to consider moving towards modern formats for solid Earth geophysical data to enable greater uptake of high resolution, data-intensive analysis in the Resource Exploration Industry, particularly as there is increasing usage of AI and ML and more open-source tools are becoming available.

For the future, to realise the full value of significant investment going back over 50 years in acquiring continental-scale geophysical datasets in government, industry and academia, we need to consider a collaborative effort from all three sectors to:

1. Remaster our current geophysical assets into modern HPC-compatible, self-describing formats suitable for in-situ processing to enable high-resolution processing at scale, and readiness for new computational techniques (e.g., Machine Learning, Artificial Intelligence);
2. Attach rich-metadata that ensure full compliance for both humans and machines with FAIR (Wilkinson et al., 2016) and CARE principles (Carroll et al., 2020) at the dataset level, and develop methods for automatically generating required-rich metadata; and
3. Ensure that the national scale (meta)data are fully machine-actionable and link with major international geophysical research infrastructures, including Geo-INQUIRE, EPOS, ChEESE, EarthScope; to enable the Australian datasets and Australian researchers to be part of global leading-edge HPC geophysics experiments.

There is an urgency in seriously considering and investing in the above proposal. As was found in the 2030 Geophysics Collections Project, due to the lack of investment in appropriate storage facilities, many important raw and minimally processed datasets are stored offline and are at risk of deterioration of the media they are stored on or even worse, complete loss as few are still around who know what is stored on which disk/external hard drive and where.

However, it is horses for courses, and for the foreseeable future there is still a large community of Australian geophysicists who will rely on existing formats and tools. As the new data-intensive high resolution geophysics data infrastructures are developed and exploited, this existing community cannot be forgotten: one size no longer fits all.

ACKNOWLEDGMENTS

We acknowledge the willing assistance and persistence of Bruce Goleby (OPM consulting), Stephan Thiel and Kate Brand (formerly of the Geological Survey of South Australia), Goran Boren and Graham Heinson (The University of Adelaide), Tom Ostersen (formerly of the University of Tasmania) and Mark Duffett (Mineral Resources Tasmania) in helping enrich the metadata, and in some cases, find the original AusLAMP time series datasets for the Musgraves, South Australia and Tasmania. Robert Pickle, Michelle Salmon and Voon Hui Lai (Australian National University) are thanked for their help in ingesting the AusPASS PS and RSES DAS datasets. This work has been financially supported by co investments from AuScope, NCI and the ARDC Cross-NCRIS National Data Assets Project XN002: National high-resolution geophysics reference collections for 2030 computation https://doi.org/10.47486/XN002 .

REFERENCES

AuScope, 2023, AuScope Distributed Acoustic Sensing (DAS) Collection, https://dx.doi.org/10.25914/zr9f-1e98

Berglund, H.T., Trabant, C., Mencin, D., Carter, J., Briggs, M., Stults, M., Szeliga, W.M., Gottlieb, M.H., Dittmann, T.T. and Bates, G., 2022, GeoCrate: Modern Data Containers For Geophysical Data in the Cloud. In: AGU Fall Meeting Abstracts (Vol. 2022, pp. IN42D-0350), https://ui.adsabs.harvard.edu/abs/2022AGUFMIN42D0350B/abstract

Carroll, S., Garba, I., Figueroa-Rodríguez, O., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R. and Sara, R., 2020, The CARE principles for indigenous data governance: Data Science Journal, 19(1), p. 43,https://doi.org/10.5334/dsj-2020-043

Cocco, M., Freda, C., Atakan, K., Bailo, D., Saleh-Contell, K., Lange, O. and Michalek, J. (2022) "The EPOS Research Infrastructure: a federated approach to integrate solid Earth science data and services", *Annals of Geophysics*, 65(2), p. DM208. https://doi/org/10.4401/ag-8756

Cotton, F., Strollo, A., Pedersen, H., Crowley, H., Wiemer, S., Haslinger, F., Urvois, M., Schmittbuhl, J., Lorito, S., Babeyko, A., Bailo, D., Michalek, J., Lange, O., Quintero, J., Festa, G., Murphy, S., Majdanski, M., Christadle, I., Prestes, M., and Weege, S., Advancing frontier knowledge of the solid earth by providing access to integrated and customized services: the Geo-INQUIRE project: EGU General Assembly 2023, Vienna, Austria, 23–28 Apr 2023, EGU23-12686, https://doi.org/10.5194/egusphere-egu23-12686

Evans, B., Wyborn, L., Croucher, J., Rees, N., Hollmann, H., and Farrington, R., 2024, Building a National High-Resolution Geophysics Reference Collection for 2030 Computation (1.0): Zenodo, https://doi.org/10.5281/zenodo.11118382

Folch, A., Abril, C., Afanasiev, M., Amati, G., Bader, M., Badia, R.M., Bayraktar, H.B., Barsotti, S., Basili, R., Bernardi, F. and Boehm, C., 2023, The EU Center of Excellence for Exascale in Solid Earth (ChEESE): Implementation, results, and roadmap for the second phase: Future Generation Computer Systems, 146, pp.47-61. https://doi.org/10.1016/j.future.2023.04.006

Folk, M., McGrath, R.E. and Yeager, N., 1999, HDF: an update and future directions. In: IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No. 99CH36293) (Vol. 1, pp. 273-275), IEEE, https://doi.org/10.1109/IGARSS.1999.773469

Gorbatov, A., Czarnota, K., Hejrani, B., Haynes, M., Hassan, R., Medlin, A., Zhao, J., Zhang, F., Salmon, M., Tkalčić, H., Yuan, H., Dentith, M., Rawlinson, N., Reading, A.M., Kennett, B.L.N., Bugden, C. and Costelloe, M., 2020, AusArray: quality passive seismic data to underpin updatable national velocity models of the lithosphere. In: Czarnota, K., Roach, I., Abbott, S., Haynes, M., Kositcin, N., Ray, A. and Slatter, E. (eds.) Exploring for the Future: Extended Abstracts, Geoscience Australia, Canberra, 1–4, http://dx.doi.org/10.11636/135284

Hassan, R., Hejrani, B., Medlin, A., Gorbatov, A. and Zhang, F., 2020, High-performance seismological tools (HiPerSeis). In: Czarnota, K., Roach, I., Abbott, S., Haynes, M., Kositcin, N., Ray, A. and Slatter, E. (eds.) Exploring for the Future: Extended Abstracts, Geoscience Australia, Canberra, 1–4. http://dx.doi.org/10.11636/135095

Hey, T. and Trefethen, A., 2003, The data deluge: An e-science perspective: Grid computing: Making the global infrastructure a reality, pp.809-824. https://doi.org/10.1002/0470867167.ch36

Hui Lai, V., Hodgkinson, K.M., Porritt, R.W. and Mellors, R., 2024, Toward a Metadata Standard for Distributed Acoustic Sensing (DAS) Data Collection: Seismological Research Letters, 95(3), pp.1986-1999. https://doi.org/10.1785/0220230325

Ip, A., Turner, A., Poudjom-Djomani, Y., Brodie, R.C., Wynne, P., Druken, K., Symington, N. and Kemp, C., 2019, Discovering and using geophysical data in the 21st century: ASEG Extended Abstracts, 2019(1), pp.1-6. https://doi.org/10.1080/22020586.2019.12073191

James, S.R., Foks, N.L. and Minsley, B.J., 2022, GSPy: A new toolbox and data standard for Geophysical Datasets: Frontiers in Earth Science, 10, p.907614. https://doi.org/10.3389/feart.2022.907614

Kelbert, A., 2020, EMTF XML: New data interchange format and conversion tools for electromagnetic transfer functions: Geophysics, 85(1), pp.F1-F17. https://doi.org/10.1190/geo2018-0679.1

Kelbert, A., Erofeeva, S., Trabant, C., Karstens, R., and Van Fossen, M., 2018, Taking magnetotelluric data out of the drawer: Eos, 99, https://doi.org/10.1029/2018EO112859.

Krischer, L., Smith, J., Lei, W., Lefebvre, M., Ruan, Y., Sales de Andrade, E., Podhorszki, N., Bozdağ, E., and Tromp, J., 2016, An Adaptable Seismic Data Format: Geophysical Journal International, Volume 207, Issue 2, November 2016, Pages 1003–1011, https://doi.org/10.1093/gji/ggw319

Ni, Y., Denolle, M.A., Fatland, R., Alterman, N., Lipovsky, B.P. and Knuth, F., 2024, An Object Storage for Distributed Acoustic Sensing: Seismological Research Letters, 95(1), pp.499-511. https://doi.org/10.1785/0220230172

Peacock, J., Kappler, K., Heagy, L., Ronan, T., Kelbert, A. and Frassetto, A., 2022, MTH5: An archive and exchangeable data format for magnetotelluric time series data: Computers and Geosciences, 162, 105102, https://doi.org/10.1016/j.cageo.2022.105102

Rees, N., Wyborn, L., Evans, B., Farrington, R., Rawling, T., Yang, R. and Sun, Y., 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

Salmon, M., Pickle, R., Allgeyer, S. and Pfeffer, J., 2020, AusPass, making seismic data stewardship FAIR. In: Abstracts for the CSIRO Sub 20 - Inaugural Deep Earth Imaging Conference, Perth, February 2020, https://wp.csiro.au/sub20/files/2020/02/Michelle_Salmon.pdf

The University of Adelaide, 2020, Magnetotelluric surveys of the Australian Lithospheric Architecture Magnetotelluric Project (AusLAMP)
[Data collection], https://dx.doi.org/10.25914/5eaa30cc934d0

The University of Adelaide, 2022, Magnetotelluric surveys of the AusLAMP Musgraves Province, 2016 to 2018 [Data set], https://dx.doi.org/10.25914/rrm2-b189

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E. and Bouwman, J., 2016, The FAIR Guiding Principles for scientific data management and stewardship: Scientific data, 3(1), pp.1-9. https://doi.org/10.1038/sdata.2016.18

Top500 the list: https://top500.org/lists/top500/1993/11/ (1993). Accessed 15 June 2024

Top500 the list: https://top500.org/lists/top500/list/2024/06/ (2024). Accessed 15 June 2024

Trabant, C., Berglund, H., Carter, J. and Mencin, D., 2022, Developing a Next Generation Platform for Geodetic, Seismological and Other Geophysical Data Sets and Services. In: EGU General Assembly Conference Abstracts (pp. EGU22-8905), https://doi.org/10.5194/egusphere-egu22-8905

Wight, D.E., 1987, Society of Exploration Geophysicists MT/EMAP Data Interchange Standard. Available at https://library.seg.org/pb-assets/technical-standards/seg_mt_emap_1987-1686080985537.pdf

Wyborn, L., and Woodcock, R., 2010, Thinking Out Cloud: the need for Increasing Virtualisation of Australian Geoscience Data to Underpin Cost Effective Deep Exploration. In: Australian Earth Sciences Convention 2010, Abstracts 98, Geological Society of Australia, Sydney, 398-399, https://doi.org/10.13140/RG.2.2.25489.03681

Wyborn, L., Rees, N., Croucher, J., Hollmann, H., Farrington, R., Evans, B., Thiel, S., Duffett, M., and Rawling, T., 2024, Who has got what where? FAIR-ly coordinating multiple levels of geophysical data products over distributed Research Infrastructures (RIs) to meet diverse computational needs and capabilities of users: EGU General Assembly 2024, Vienna, Austria, 14–19 Apr 2024, EGU24-14052, https://doi.org/10.5194/egusphere-egu24-14052

Zawacki, E. E., Bendick, R., and Woodward, R. L., 2023, Advancing geophysics: IRIS and UNAVCO merge to form EarthScope Consortium. Perspectives of Earth and Space Scientists, 4, e2023CN000227, https://doi.org/10.1029/2023CN000227