Geophysical Research Abstracts
Vol. 18, EGU2016-5578, 2016
EGU General Assembly 2016

NCI's Transdisciplinary High Performance Scientific Data Platform

Ben Evans (1), Joseph Antony (1), Irina Bastrakova (2), Nicholas Car (2), Simon Cox (3), Kelsey Druken (1), Bradley Evans (4), Ryan Fraser (3), Alex Ip (2), Carina Kemp (2), Edward King (3), Stuart Minchin (2), Pablo Larraondo (1), Tim Pugh (5), Clare Richards (1), Fabiana Santana (1), Jon Smillie (1), Claire Trenham (1), Jingbo Wang (1), and Lesley Wyborn (1)

(1) National Computational Infrastructure, Australian National University, Canberra, Australia (ben.evans@anu.edu.au), (2) Geoscience Australia, Symonston, Australia, (3) Commonwealth Scientific and Industrial Research Organisation, CSIRO, Australia, (4) Sydney University, Sydney, Australia, (5) The Bureau of Meteorology, Melbourne, Australia

The Australian National Computational Infrastructure (NCI) manages Earth Systems data collections sourced from several domains and organisations onto a single High Performance Data (HPD) Node to further Australia's national priority research and innovation agenda. The NCI HPD Node has rapidly established its value, currently managing over 10 PBytes of datasets from collections that span a wide range of disciplines including climate, weather, environment, geoscience, geophysics, water resources and social sciences. Importantly, in order to facilitate broad user uptake, maximise reuse and enable transdisciplinary access through software and standardised interfaces, the datasets, associated information systems and processes have been incorporated into the design and operation of a unified platform that NCI has called, the National Environmental Research Data Interoperability Platform (NERDIP).

The key goal of the NERDIP is to regularise data access so that it is easily discoverable, interoperable for different domains and enabled for high performance methods. It adopts and implements international standards and data conventions, and promotes scientific integrity within a high performance computing and data analysis environment. NCI has established a rich and flexible computing environment to access to this data, through the NCI supercomputer; a private cloud that supports both domain focused virtual laboratories and in-common interactive analysis interfaces; as well as remotely through scalable data services.

Data collections of this importance must be managed with careful consideration of both their current use and the needs of the end-communities, as well as its future potential use, such as transitioning to more advanced software and improved methods. It is therefore critical that the data platform is both well-managed and trusted for stable production use (including transparency and reproducibility), agile enough to incorporate new technological advances and additional communities practices, and a foundation for new exploratory developments. To that end, NCI is already participating in numerous current and emerging collaborations internationally including the Earth System Grid Federation (ESGF); Climate and Weather Data from International agencies such as NASA, NOAA, and UK Met Office; Remotely Sensed Satellite Earth Imaging through collaborations through GEOS and CEOS; EU-led Ocean Data Interoperability Platform (ODIP) and Horizon2020 Earth Server2 project; as well as broader data infrastructure community activities such as Research Data Alliance (RDA).

Each research community is heavily engaged in international standards such as ISO, OGC and W3C, adopting community-led conventions for data, supporting improved data organisation such as controlled vocabularies, and creating workflows that use mature APIs and data services. NCI is engaging with these communities on NERDIP to ensure that such standards are applied uniformly and tested in practice by working with the variety of data and technologies. This includes benchmarking exemplar cases from individual communities, documenting their use of standards, and evaluating their practical use of the different technologies. Such a process fully establishes the functionality and performance, and is required to safely transition when improvements or rationalisation is required.
Work is now underway to extend the NERDIP platform for better utilisation in the subsurface geophysical community, including maximizing national uptake, as well as better integration with international science platforms.