Geophysical Research Abstracts
Vol. 20, EGU2018-12639, 2018
EGU General Assembly 2018


# First steps towards internationally integrating data and services in the solid Earth sciences and beyond

Lesley Wyborn (1), Ben Evans (2), Kerstin Lehnert (2), Tim Rawling (3), Jens Klump (4), Kirsten Elger (5), SImon Cox (6), Helen Glaves (7), Mohan Ramamurthy (8), Erin Robinson (9), and Shelley Stall (10)

(1) Australian National University, National Computational Infrastructure, Acton, Australia (lesley.wyborn@anu.edu.au, ben.evans@anu.edu.au), (2) Lamont-Doherty Earth Observatory, Columbvia University, New York, USA (lehnert@ldeo.columbia.edu), (3) AuScope Ltd, Melbourne, Australia (tim.rawling@unimelb.edu.au), (4) Mineral Resources, CSIRO, Kensington, WA, Australia (jens.klump@csiro.au), (5) GFZ German Research Centre for Geosciences, Potsdam, Germany (kelger@gfz-potsdam.de), (6) Land and Water, CSIRO, Clayton, Vic, Australia (simon.cox@csiro.au), (7) British Geological Survey, Nottingham, UK (hmg@bgs.ac.uk), (8) EarthCube, UCAR, Boulder, USA (mohan@ucar.edu), (9) Earth Science Information Partners, Boulder, CO, United States (erinrobinson@esipfed.org), (10) American Geophysical Union, Washington, D.C., United States (sstall@agu.org)

Globally, solid Earth science data are collected by large numbers of organizations across the academic, government and industry sectors. Spatially, the data collected covers multiple domains extending from the crust, through the lithosphere and mantle to the core. In all, many observed phenomena cross national, if not continental, boundaries, and increasingly require international networks of researchers to address growing global challenges such as scarce non-renewable resources, risk reduction for natural hazards, and fundamental research on the nature of the planet.

The last decade has seen a dramatic growth in the number of online solid Earth science datasets and in online computational power, particularly utilising Cloud or HPC hosted data and compute resources. However, data in many of these online resources have inconsistent and incompatible data descriptions and formats, and as much as 80% of data processing effort is spent on discovering, cleaning and converting pre-existing data. Software are often developed locally around specific applications and data sources, with the side-effect of a multiplicity of software providing similar and overlapping functions.

To be able to address growing global research challenges, more attention needs to be paid to harmonising multiple metadata/data standards to enable globally discoverable and accessible data, and to enhancing standards that make data programmatically actionable from robust data platforms. Knowing that data and the data access meets agreed standards means that the software community can focus on developing better algorithms, rather than creating a myriad of ways of accessing the same data type in multiple formats. It will also be easier to create workflows for those outside of the more specialised research community.

Currently there are established national efforts creating infrastructures that help connect solid Earth researchers. In the US these include the Earth Science Information Partners (ESIP) and the NSF's EarthCube program, whilst in Australia there has been rapid advancements in supporting e-Infrastructure with investments such as AuScope and the National Computational Infrastructure (NCI). In Europe equivalent Horizon2020 projects are the Environmental Research Infrastructure Plus (ENVRIplus) and European Plate Observing System (EPOS). All are linking data, cyberinfrastructure and research developments across the academic and government sectors. Furthermore, more generic software from standards bodies now support many of the core requirements directly (e.g., W3C's DCAT metadata vocabulary, W3C/OGC's SOSA/SSN observations and sampling ontology, DataCite identifier and metadata systems).

What is needed now are mechanisms to internationally link these major infrastructures to provide not only efficiencies in funding, but also an environment where the research efforts can create globally interoperable networks of solid Earth science data, information systems, software and researchers. Furthermore, the solid Earth community also needs to understand how to build its data networks to be compatible with those of other communities. Data of similar forms are being collected 'above Earth' in the atmosphere, biosphere, cryosphere, hydrosphere, and pedosphere. To prepare for the future transdisciplinary science challenges, these data will be more valuable when linked with equivalent activities in data and services for environmental, atmospheric, climate and marine research.