Geophysical Research Abstracts
Vol. 20, EGU2018-14055, 2018
EGU General Assembly 2018


# Data Infrastructure for the Earth & Space Sciences: How Far Have We Come, Where Are We Heading?

Kerstin Lehnert
Columbia University, Lamont-Doherty Earth Observatory, Palisades, United States (lehnert@ldeo.columbia.edu)

Over the past 10 years, the global ecosystem of research data has evolved dramatically. New technologies for data storage, description, identification, interoperability, and analytics have emerged that considerably advance researchers' ability to discover and access data across distributed systems, and then re-use them to answer new science questions. New policies for data management planning and open data sharing have been developed and put in place by funders and publishers in many countries to ensure preservation and FAIRness of data (Findable, Accessible, Interoperable, Reusable). And new initiatives and organizations have come to life such as the Research Data Alliance and the World Data System within the past decade to advance cross-disciplinary and international communication, coordination and, collaboration, helping to build consensus on best practices and standards for the sharing and re-using research data.

The Earth and space science (ESS) community has been an active participant and a driving force in this technical, organizational, and social transformation. The availability of data and data services for the ESS community has made substantial progress. But how far along are we in the process of building a persistent and reliable infrastructure for our ESS research data? How comprehensive and inclusive are the current capabilities and services, how well do they serve the full spectrum of disciplinary communities from the big head to the long tail? Have we been able to get all stakeholders in the research data ecosystem on board and create a new community of practice? How sustainable are the systems and services that are now in place, are funders stepping up to provide the needed long term support for data infrastructure, or have alternative business models been found? And most importantly, are data infrastructures advancing science and how do we assess that?

This lecture will examine the past decade of developments in the Earth and space science data ecosystem, mapping progress and challenges to known patterns of infrastructure development as described by Paul Edwards and co-authors in their report on "Understanding Infrastructure: Dynamics, Tensions, Design", published in 2007. Ten years ago, Edward's report generated awareness that the most significant barriers that impede infrastructure development are not technical, but legal, political, social, or cultural in nature. In 2008, I used the insights of Paul Edward's report to highlight progress of data infrastructure for geochemistry and understand barriers to adoption of data sharing practices in this long tail community. In this lecture I will explore how the ESS research data infrastructure has evolved, comparing current status to the one ten years ago. Which barriers have we overcome over the past decade and which ones still remain to be eliminated? For example, initiatives such as COPDESS and the AGU FAIR project that advance consistent implementation of leading practices for data in scholarly publications are changing the data sharing culture in the long tail. A fundamental problem remains the lack of long-term, stable support for domain data repositories that provide the relevant high-quality curation services for the community to make data FAIR.