



Sharing Infrastructure Among Data Facilities: An Initiative of the EarthCube Council of Data Facilities

Kerstin Lehnert (1), Timothy Ahern (2), and Danie Kinkade (3)

(1) Columbia University, Lamont-Doherty Earth Observatory, Palisades, United States (lehnert@ldeo.columbia.edu), (2) IRIS Data Management Center Seattle, WA , United States, (3) Woods Hole Oceanographic Institution, Woods Hole, MA, United States

Data facilities in the Earth, Space, and Environmental sciences provide indispensable services to the science community ensuring discovery, access, reusability, attribution, and preservation of data as essential products of and resources for scientific research. Most data facilities these days are challenged with meeting the growing demands of users and funders, and of a maturing research data ecosystem, from the rapidly expanding volumes of data that the facilities need to manage, requiring new levels of storage and network capacity and, consequentially, resources and team expertise; to the need to comply with international guidelines for trustworthy operation of services, including transparent and standards-based data curation procedures, licensing, risk management, security, and sustainability; to continuously evolving opportunities or requirements for enhancing machine access and interoperability of data holdings; to the need to maintain a diverse staff that keeps up with new and emerging technologies. In addition, they struggle with stagnant or declining funding and longer-term sustainability. Establishing infrastructure services that are shared among multiple data facilities can potentially help each facility better address these challenges, making operations more efficient and sustainable, while also leading to better alignment of technologies, policies, and procedures across data facilities, and opening opportunities for joint developments and innovation to meet future needs.

The Council of Data Facilities (CDF) is a federation of existing and emerging geoscience data facilities that serves as an effective foundation for the US NSF EarthCube program and related contributors to delivering cyberinfrastructure for earth system science. One of the declared goals of the CDF since its inception has been to identify and support the development and utilization of shared infrastructure services, including computing and storage services, professional staff development and training services, and related activities. In 2018, a CDF Working Group was established to explore priorities and first steps for developing shared infrastructure. A preliminary survey was completed by 11 data facilities in 2018 and identified as priorities for shared infrastructure: cloud services (VMs, storage, computing, shared licenses); shared subscriptions for DOI minting and access to ORCID services; access to a shared pool of technical capacity (expertise & person hours) as well as shared expertise for repository certification; and joint training and knowledge sharing mechanisms. A more detailed assessment of facilities' needs for shared infrastructure is currently underway and will be completed in January 2019. This presentation will report on the results of this assessment and the CDF's plans for the next steps toward implementation of shared infrastructure services.