Geophysical Research Abstracts
Vol. 15, EGU2013-13218, 2013
EGU General Assembly 2013

Data Publication: The Role of Community-Based, Disciplinary Repositories

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

Data publication is increasingly recognized as a new paradigm to achieve broad and open access to data for their re-use while ensuring proper citation and attribution of the data to their creator. Current models of data publication include special data journals, institutional repositories, and programmatic or disciplinary repositories.

This presentation focuses on the critical role that community-based disciplinary repositories should play, using the example of the IEDA data facility (www.iedadata.org). Disciplinary repositories ensure that data publication is successfully implemented and adopted by the science community, so that it can fulfill its promise to advance data reuse, enhance reproducibility of scientific results, and offer new opportunities for scientific discoveries. Disciplinary repositories are ideally poised to address the challenges that can lead to 'data reuse failure' (Rees 2010), which in turn obstruct the successful development and implementation of new data infrastructures for the sciences. Among these challenges are: professional motivation for investigators to publish the data; effort and economic burden of publication; discovery and long-term accessibility of data; adequate documentation of data provenance (measured parameters, uncertainty, reproducibility, materials, methods, data manipulation); and usability (file formats, interfaces to software tools for data analysis, visualization, modeling).

The relevant function of a disciplinary repository is to coordinate and bridge the needs of its community with recognized requirements for data curation and demands of evolving technologies. IEDA meets that function, gathering regular guidance from the community, developing community-approved best practices and guidelines; providing guidance and tools to investigators that facilitate the process of data publication; and performing data quality control. IEDA's sustainable infrastructure supports the discovery and access of the data, and IEDA continues to develop advanced tools for capture of data and metadata as part of the entire workflow form data acquisition to data reduction and data analysis.

As a result of IEDA's community-based approach, IEDA has increased publication data that may otherwise have become dark data or not been sufficiently documented for reuse.

Reference:
Rees, Jonathan (2010): "Recommendations for independent scholarly publication of data sets." Creative Commons Working Paper