Empowering Earth Science Communities to Share Data Through Guided Metadata Improvement: DataONE Webinar

Lindsay Powers, The HDF Group

The HDF Group is collaborating with the University of California, Santa Barbara and Data Observation Network for Earth (DataONE), to help scientific research communities enhance the consistency and quality of their metadata, and to foster discovery, access and understanding of data resources. As part of this collaboration, on February 9, 2016, The HDF Group's Ted Habermann, Director of Earth Science, and Lindsay Powers, Deputy Director of Earth Science will co-lead a webinar, "Sharing Data Through Guided Metadata Improvement" along with Matthew Jones, Director of Informatics Research at the National Center for Ecological Analysis and Synthesis.

DataONE is a distributed network of data centers, science networks or organizations that comprise the Member Nodes (scientific communities). The network provides integrated access to the scientific data holdings of the Member Nodes through a centralized cyber-infrastructure that enables access to data and customized tools, and provides community engagement and outreach designed to enable new science and knowledge creation. The DataONE Webinar Series is designed to engage participants in relevant and cutting-edge topics within the Earth and environmental sciences. The series features discussions on open science, the role of the data lifecycle, and achieving innovative science through shared data and groundbreaking tools.

In this webinar we will describe the work that is ongoing with DataONE Member Nodes to quantitatively evaluate their metadata and to identify specific strategies to improve the completeness and consistency of their metadata. We will discuss the preliminary results addressing three hypotheses:

- Can community-developed metadata recommendations help improve metadata content within a particular community?
- Can community-developed metadata recommendations help improve metadata content among communities?
- Can metadata recommendations developed in a specific dialect be used to help improve metadata in other dialects?

We have developed an iterative, guided process intended to efficiently improve metadata to better serve individual communities, as well as share data across disciplines. The community-specific approach focuses on community metadata requirements, and also provides guidance on adding other metadata concepts to expand the effectiveness of metadata for multiple uses, including data discovery, data understanding, and data re-use. The end goal of this work is to help communities improve their metadata based on their own requirements through time.

We will present the results of a baseline analysis of more than 15 diverse metadata collections from established data repositories representing communities across the earth and environmental sciences. The baseline analysis describes the current state of the metadata in these collections and highlights areas for improvement. We compare these collections to demonstrate exemplar practitioners that can provide guidance to other communities.

In addition, we are building web-based tools founded on a common metadata evaluation library that can be incorporated into community tools such as metadata editors and repository platforms, as well as form the core of a metadata completeness reporting service that is integrated within specific partner information systems such as the DataONE Coordinating Node services and the Mercury Online Metadata Editor. This aspect of the project is forthcoming and we will discuss the plans for the future.

DataONE Webinar information:

- Webinars will occur monthly at 9 AM US Pacific time (12PM US Eastern time) on the second Tuesday of the month, starting February 10th 2015. Each webinar will last one hour, with 30-40 minutes for the presentation and 20-30 minutes for discussion

- Remote access using GoToWebinar

- Webinars will be recorded and posted on this website afterward

- An open forum will be available for 24 hours after the webinar to facilitate further discussion between participants and the speaker

- Check out the DataONE Google Calendar for updates

https://www.dataone.org/upcoming-webinar

Register for the webinar at https://attendee.gotowebinar.com/register/988575674277326594

* * * * *

The HDF5 data model, file format, API, library, and tools are open and distributed without charge.

Building on its 28-year history, The HDF Group offers personalized consulting, training, design, software development, and support services to help clients take full advantage of HDF5 capabilities in addressing their unique data management challenges. To discuss these services, please contact the HDF Group Help Desk (help@hdfgroup.org).

The newest HDF5-1.10.0-alpha release can be obtained from The HDF Group Downloads page: https://www.hdfgroup.org/downloads/

It can also be obtained directly from: https://www.hdfgroup.org/HDF5/release/obtain5110.html

Documentation can be found in: https://www.hdfgroup.org/HDF5/docNewFeatures/