U.S. DEPARTMENT OF ENERGY | Office of Science

CESD Cyberinfrastructure Working Groups

Environmental System Science (ESS) PI Meeting

Bolger Center, Potomac, Maryland, USA

April 30, 2018

Data Management

Lead: Charuleka Varadharajan (LBNL)

Current Members: Deb Agarwal, Danielle Christianson, Paul Hanson, Val Hendrix, Trevor Keenan, Terri Velliquette, Lara Kueppers, Dave Millard, Eric Pierce, Gilberto Pastorello, Daniel Ricciuto, Cory Snavely, James Stegen, Margaret Torn, Roelof Versteeg, Ken Williams, Chongang Xu

U.S. DEPARTMENT OF ENERGY | Office of Science

What is ESS Data?

“Data” = Field Observations, Experimental Data, Remote Sensing Data, Analysis and synthesis products, Model data, data in publications

Scope of the Data Working Group Activities

- **Management and Archival** of DOE climate and environmental datasets
  - –Data **Preservation, Sharing, and Publication**
  - –Common Data and Metadata **Standards**
  - –Data **Citation and Attribution**
  - –Data **Federation** across different data catalogs
- Data **Synthesis** across ESS and other relevant Datasets
- Development of common **Tools** for data usage
- QA/QC, processing, analysis, mining and visualization data to prepare them for use in new research projects.

ESS-DIVE: A New Data Archive for Earth and Environmental Science Data

http://ess-dive.lbl.gov

U.S. DEPARTMENT OF ENERGY | Office of Science

CESD Data Management Program

BERKELEY LAB

NCEAS

NERSC

DataONE
Data Observation Network for Earth

To **preserve**, expand **access** to, and improve **usability** of critical **data** generated through DOE-sponsored research **of terrestrial and subsurface ecosystems** in support of the DOE's efforts to address some of society's most pressing energy and environmental challenges.

Support Data Infrastructure

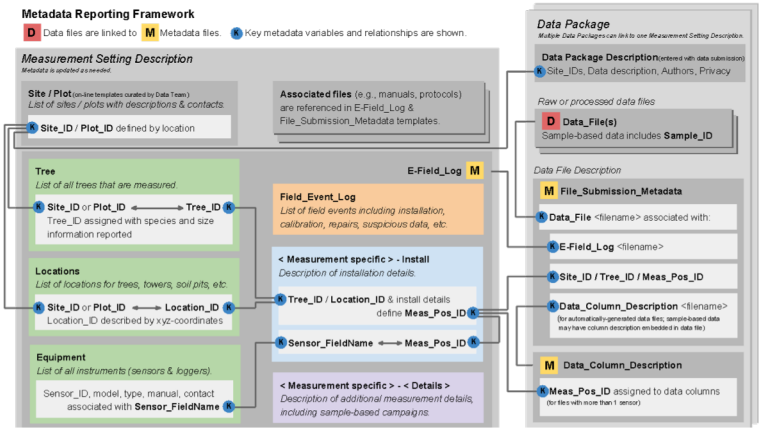
- Working group supports CESD/ESS data infrastructure with input from community
- WG members participated in defining proposal call for the ESS data archive based on identified data management needs
  - –Common data repository across ESS with federated logins
  - –Metadata development and editing tools
  - –Unified procedures for data publishing, archiving, and federation
  - –Established data policies and data citation/attribution mechanisms
- Provided input into ESS-DIVE initial package metadata

Short term goals (2015-2017)

- Shared database for published data
  - Develop workflow to capture and serve publication data that meets DOE digital data requirements; define publication data packages
  - Define a common standard for using DOI or similar digital identifiers to uniquely identify electronic data for publication and citation
  - Best practices and use case driven templates for data archiving beginning with observational data, and extending to all supporting data, workflow information
  - Centralized DOE single sign-on
- Identify Best Practices for Metadata
- Sample Tracking

Short term goals for the upcoming year

- Continue to engage and provide feedback in ESS-DIVE design
- Work on metadata standards for ESS data to archived data usable
  - –Identifying relevant existing standards (e.g. netCDF, EML)
  - –Defining new standards where needed (e.g. geochemical, spectroscopic data)
- Sample tracking, standardized reporting of sample
- Tools for data QA/QC, analysis, synthesis, visualizations in conjunction with science needs



Long-Term Goals