

Poster #BF-Hall

Environmental System Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE) - A New U.S. DOE Data Archive

Deb Agarwal^{1*}, Charuleka Varadharajan¹, Shreyas Cholia^{1,2}, Cory Snavely², Valerie Hendrix¹, Fianna O'Brien¹, Abdelrahman Elbashandy¹, Dan Gunter¹, William Riley¹, Chris Jones³, Matt Jones³, Amber E Budden⁴, and Dave Vieglais⁵

¹Lawrence Berkeley National Laboratory, Berkeley, CA

²NERSC, Berkeley, CA

³NCEAS, Santa Barbara, CA

⁴DataONE, Santa Barbara, CA

⁵University of Kansas, Lawrence, KS

Contact: daagarwal@lbl.gov

BER Program: CESD Data Management

Project: ESS-DIVE

Project Website: <http://ess-dive.lbl.gov>

The ESS-DIVE archive is a new U.S. Department of Energy (DOE) data archive designed to provide long-term stewardship and use of data from observational, experimental, and modeling activities in the earth and environmental sciences. The ESS-DIVE infrastructure is constructed with the long-term vision of enabling broad access to and usage of the DOE sponsored data stored in the archive. It is designed as a scalable framework that incentivizes data providers to contribute well-structured, high-quality data to the archive and that enables the user community to easily build data processing, synthesis, and analysis capabilities using those data.

The key innovations in our design include: (1) application of user-experience research methods to understand the needs of users and data contributors; (2) support for early data archiving during project data QA/QC and before public release; (3) focus on implementation of data standards in collaboration with the community; (4) support for community built tools for data search, interpretation, analysis, and visualization tools; (5) data fusion database to support search of the data extracted from packages submitted and data available in partner data systems such as the Earth System Grid Federation (ESGF); and (6) support for archiving of data packages that are not to be released to the public.

ESS-DIVE data contributors will be able to archive and version their data and metadata, obtain data DOIs, search for and access ESS data and metadata via web and programmatic portals, and provide data and metadata in standardized forms. The ESS-DIVE archive and catalog will be federated with other existing catalogs, allowing cross-catalog metadata search and data exchange with existing systems. ESS-DIVE is operated by a multidisciplinary team from Berkeley Lab, the National Center for Ecological Analysis and Synthesis (NCEAS), and DataONE. The primary data copies are hosted at DOE's NERSC supercomputing facility with replicas at DataONE nodes.