

Poster #21-30

Data Management Framework for LBNL's Watershed Function Scientific Focus Area

Charuleka Varadharajan^{1*}, Danielle Christianson¹, Dipankar Dwivedi¹, Boris Faybishenko¹, Matthew Henderson¹, Valerie Hendrix¹, Zarine Kakalia¹, Roelof Versteeg², and Deb Agarwal¹

¹Lawrence Berkeley National Laboratory, Berkeley, CA

²Subsurface Insights

Contact: cvaradharajan@lbl.gov

BER Program: SBR

Project: Berkeley Lab Watershed Function SFA

Project Website: watershed.lbl.gov

The Watershed Function SFA project generates diverse datasets at its field site in East River, Colorado, which include hydrological, geochemical, geophysical, ecological, microbiological, and remote sensing data. The datasets are collected by the LBNL project team and obtained from collaborators or other external sources.

The objective of the SFA's Data Management Framework is to provide infrastructure and services to: (1) manage, archive, and publicly release data collected by the SFA as per the [project's data policy](#); (2) enable the SFA team and the broader community to discover and access relevant datasets; (3) perform Quality Assurance and Quality Control (QA/QC) of priority datasets; and (4) enable efficient data collection, data integration, and product generation.

To meet these objectives, the data team has developed a number of tools for data management and preservation, QA/QC, data discovery, advanced search, and visualization. The SFA's data package system provides for internal data archival and curation of project-generated data. It allows authorized users to upload data files with relevant metadata, and download files shared within the project team. The SFA's datasets are publicly released via the ESS-DIVE archive. Priority datasets are run through automated QA/QC algorithms that are based on the application of statistical methods used to identify and flag bad and anomalous data in collected datasets. The data team is developing a brokering service BASIN-3D that integrates diverse project and external datasets in real-time from distributed databases via web services. BASIN-3D is based on the Open Geospatial Consortium's Observations and Measurements standard and is currently implemented to integrate time-series data. It powers a [web portal](#) that enables users to conduct an intuitive search, interactive visualization, and download of integrated datasets. Additionally, a [public interactive map](#) of the East River has been developed to inform the broader community about the SFA's field infrastructure and research activities. Sites can be filtered by their key measurements and other metadata, leading to detailed site landing pages.

The data team is working in cooperation with the science teams to create crosscutting data products and workflows for hypothesis testing and modeling. Standardized field collection sheets and sample tracking were implemented to coordinate the extensive data and sample collection for the NEON hyperspectral ground campaign. The developed QA/QC'ed meteorological data products are being used by the science teams as model inputs for predictions of the water balance processes, including infiltration, evapotranspiration, groundwater recharge, and river discharge, as well as geochemical modeling.