

## **An End-End Pipeline for Watershed Data Management and Assimilation**

Charuleka Varadharajan<sup>1,\*</sup>, Roelof Versteeg<sup>2</sup>, Boris Faybishenko<sup>1</sup>, Val Hendrix<sup>1</sup>, Danielle Christianson<sup>1</sup>, Dipankar Dwivedi<sup>1</sup>, Matthew Henderson<sup>1</sup>, Zarine Kakalia<sup>1</sup>, Madison Burrus<sup>1</sup>, and Deb Agarwal<sup>1</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory, Berkeley, CA;

<sup>2</sup>Subsurface Insights, Hanover, NH

**Contact:** [cvaradharajan@lbl.gov](mailto:cvaradharajan@lbl.gov)

**Project Lead Principle Investigator (PI):** Susan Hubbard

**BER Program:** SBR

**Project:** Berkeley Lab Watershed Function SFA

**Project Website:** [watershed.lbl.gov](http://watershed.lbl.gov)

### **Project Abstract:**

The Watershed Function SFA project generates heterogeneous datasets at its East River, Colorado experimental watershed that include a variety of data types, such as hydrological, geochemical, geophysical, microbiological, and remote sensing data. Data are collected from various sources, including data generated by the project team (e.g. sensor data, geochemical sampling and remote sensing products), its collaborators, and external sources (e.g. USGS and NRCS).

The Data Management Framework for the SFA provides infrastructure and services to support various aspects of the project's data lifecycle. The framework contains (a) a data collection and acquisition system involving a distributed sensor network across the watershed for diverse observations, (b) a queryable database with a workflow to telemeter and store data and associated data products with relevant metadata, (c) scripts for semi-automated QA/QC with cleaned data stored in the ERDB, (d) a data integration broker (BASIN-3D) to synthesize project data with external datasets for real-time queries, (e) an advanced data search and access portal for data discovery, exploratory analysis and download, and (f) periodic publication of data with DOIs in the DOE's ESS-DIVE repository.

These tools are used for building crosscutting data products needed for hypothesis testing and numerical modeling of hydrological and biogeochemical conditions in the East River watershed by both internal and external SFA project teams. The development and maintenance of this infrastructure presents a suite of challenges from practical field logistics to complex data processing that are addressed through various solutions. In particular, the SFA adopts a holistic view for data collection, assessment, and integration that dramatically improves the products generated, thereby enabling a co-design approach whereby data collection is informed by model results and vice-versa.