Sustainable Systems SFA Data Management and Assimilation

Charuleka Varadharajan (cvaradharajan@lbl.gov) - Lawrence Berkeley National Laboratory, Deborah Agarwal, Lawrence Berkeley National Laboratory (PI), Roelof Versteeg, Subsurface Insights; Boris Faybishenko, Lawrence Berkeley National Laboratory (Co-PIs).

A software platform and infrastructure is being developed for management and assimilation of data collected for the Sustainable Systems SFA 2.0, with components and capabilities based on the SFA 2.0 project priorities. The completed infrastructure will enable integration and synthesis of diverse and disparate field, laboratory, and simulation datasets, including geological, geochemical, geophysical, microbiological, hydrological, and meteorological data across a range of spatial and temporal scales. The current focus is to integrate monitoring data and characterization of samples collected at the field site in Rifle, Colorado.

The main elements of the Data Management System that have currently been developed include:

- RifleDB: The Rifle Data Management Web Site (http://www.rifleifrc.org/), which is used to host data from the field site. Users with a login to the website can view several datasets, including hydrological data (water levels, water and soil temperatures), geochemical data from groundwater samples, borehole lithological logs, and site meteorological data. The Data Management System also incorporates a data quality process, which is necessary to address user needs and produce quality-checked datasets. The website is implemented as a Zend Framework 2 web application, which sits on top of a MySQL database.

- SFA Google Drive: An internal Google Drive folder that is used for preliminary storage and sharing of data across all SFA components before migrating the data to a database.

- SFA Data Needs Sheet: A Smartsheet that is used to enable tracking of data that are being generated, including data owners and status. Data needs are specified for three domains - the saturated zone, vadose (unsaturated) zone, and general site information. Data items in each of the domains are categorized by a property type - e.g. hydrology, biogeochemistry, microbiology, sediment properties.

Additionally, use case scenarios are being developed to provide input into the development of products that will enable data fusion across the different sources. An example use case highlights integration of depth profile data from the vadose and saturated zones across the TT transect at the Rifle site, including geochemical measurements (concentrations of ions, TIC/TOC, and isotopes in water samples), hydrological measurements and parameters (water level, temperature, soil moisture, hydraulic conductivity and water retention curves), microbiological data (relative abundance of microorganisms and metagenomics), lithological data (borehole logs and particle size distribution), and geophysical data (time-lapse electrical resistivity tomography).