Genomes to Watershed Data Management and Assimilation

A software infrastructure and platform is being developed for data management and assimilation [DMA] as part of the LBNL Genomes to Watershed SFA 2.0, with components and capabilities driven by the project science priorities. The overall objective of the DMA software platform and infrastructure is to enable the users to integrate and synthesize 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 of the DMA effort is to (a) develop an integrated interface to the data (brokering and portal system), (b) store field monitoring data, laboratory analytical results of water and sediments samples, and metadata collected at the Rifle, Colorado, field site into a database, and (c) provide basic automated QA/QC analysis of stored data and work with data providers to modify field and laboratory data collection and reporting procedures as needed.

We have developed, tested and released a first version of our broker system that integrates diverse datasets from the RifleDB and ggkBase. This data includes hydrological parameters (water level and water temperature), geochemical analyses and metagenomic data. We have deployed a beta version of the user interface portal for the broker, based on an extensive user-centered design process. This initial version focuses on helping users find the data needed to complete analyses through an intuitive interface. Once the data is found, the user can immediately plot and download data through the portal. The interface is based on an interactive visualization, created using D3.js.

The RifleDB continues to be our primary database for storage of non-metagenomic data collected at the Rifle site and we continue to expand and improve the database. We implemented and validated automated QA/QC analyses of high-temporal resolution water level data collected by 29 pressure transducers in monitoring wells at the Rifle site. We also implemented a QA/QC analyses for vadose zone and groundwater temperature and geochemical data collected in monitoring wells, as well as meteorological data. We have provided recommendations related to the standardization of the field water-level measurements, water sampling, and reporting of the results of field and laboratory measurements.

The future goals of the DMA team include further development and testing of the brokering system, ingestion of the field and laboratory data to enable reliable integration and synthesis of diverse and disparate field, laboratory, and simulation datasets across different spatial, and temporal scales.