

EGU22-13338

<https://doi.org/10.5194/egusphere-egu22-13338>

EGU General Assembly 2022

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



## A workflow to standardize collection and management of large-scale data and metadata from environmental observatories

**Dylan O’Ryan**<sup>1,2</sup>, Charuleka Varadharajan<sup>1</sup>, Erek Alper<sup>3</sup>, Kristin Boye<sup>4</sup>, Madison Burrus<sup>1</sup>, Danielle Christianson<sup>1</sup>, Shreyas Cholia<sup>1</sup>, Robert Crystal-Ornelas<sup>1</sup>, Joan Damerow<sup>1</sup>, Wenming Dong<sup>1</sup>, Hesham Elbashandy<sup>1</sup>, Boris Faybishenko<sup>1</sup>, Valerie Hendrix<sup>1</sup>, Douglas Johnson<sup>3</sup>, Zarine Kakalia<sup>1,5</sup>, Roelof Versteeg<sup>3</sup>, Kenneth Williams<sup>1</sup>, Catherine Wong<sup>1,5</sup>, and Deborah Agarwal<sup>1</sup>

<sup>1</sup>Lawrence Berkeley National Laboratory

<sup>2</sup>California State University, Sacramento

<sup>3</sup>Subsurface Insights

<sup>4</sup>SLAC National Accelerator Laboratory

<sup>5</sup>University of California, Berkeley

The Watershed Function Scientific Focus Area (WFSFA) is a U.S. Department of Energy research project that seeks to determine how mountainous watersheds retain and release water, carbon, nutrients, and metals. The WFSFA maintains a community field observatory at its primary field site in the East River, Colorado. The WFSFA collects diverse environmental data and has developed a “Field-Data” workflow that standardizes data management across the project, from field collection to laboratory analysis to publication. This workflow enables the WFSFA to address data quality and management challenges that environmental observatories face.

Through this workflow, the WFSFA has increased the use of the data curated from the project by (1) providing detailed metadata with unique identifiers for samples, locations, and sensors, (2) streamlining the data sharing and publication process through early sharing of data internally within the team and publication of data on the ESS-DIVE repository following curation, and (3) adopting machine-readable and FAIR community data standards (Findability, Accessibility, Interoperability, Reusability).

We describe an example application of this workflow for geochemical data, which utilizes a community geochemical data standard for water-soil-sediment chemistry (<https://github.com/ess-dive-community/essdive-water-soil-sed-chem>) developed by Environmental Systems Science Data Infrastructure for a Virtual Ecosystem (ESS-DIVE). This data standard is designed to standardize geochemical data, metadata, and file-level metadata, and was applied to WFSFA geochemical data, including ICP-MS, Isotope, Ammonia-N, Anion, DIC/NPOC/TDN datasets. This ensures important metadata is contained within the data file, such as precision of data analysis, storage and sample processing information, detailed sample names, material information, and unique identifiers associated with the samples (IGSNs). This metadata is essential to understand and reuse data products, as well as enable machine-readability for future model applications. Detailed examples

of the standardized geochemical data types were created and are now being used as templates by WFSFA researchers to standardize their geochemical data. The adoption of this community geochemical data standard and more broadly the Field-Data workflow will improve the findability and reusability of WFSFA datasets.