Improving the management and reuse of water quality data for the DOE's Watershed Function Scientific Focus Area using community data standards

Dylan O'Ryan[1], Robert Crystal-Ornelas[2], Charuleka Varadharajan[2]
[1]San Joaquin Delta College, [2]Lawrence Berkeley National Laboratory

ABSTRACT

FAIR and standardized data are increasingly needed and prioritized in earth and environmental science research. In an effort toward more standardized data, Lawrence Berkeley National Lab (LBNL) has begun developing community data standards, which provide guidelines for how data providers collect and store their data. During this internship, I gained experience in data management and assisted with several essential data standardization tasks. From these Community Partner-developed water quality reporting formats (community data standards), I created examples of water quality datasets in these reporting formats to demonstrate how data providers can follow these formats when submitting data to the Environmental Systems Science Data Infrastructure for a Virtual Ecosystem data repository (ESS-DIVE). Progress that I made during the internship term helps to ensure water quality data submitted to ESS-DIVE follows Findable Accessible Interoperable and Reusable (FAIR) data practices. Future steps of community data standards will have to be integrated into current steps for data collection and ensure that necessary metadata is supplied. Additionally, LBNL has made steps to collect metadata information for sensors as part of the Watershed Function Scientific Focus Area (WFSFA). I worked with sensor metadata templates, to facilitate collecting information about sensors that are deployed as part of the WFSFA. Ultimately, data standards and metadata information will ensure that LBNL, ESS-DIVE, and the Department of Energy (DOE) follow FAIR standards.

RESEARCH QUESTION

How can a water quality reporting format being developed by ESS-DIVE Community Partners be applied to open-source water quality datasets?

BACKGROUND INFO

Terms

- reporting formats/community data standards: researcher developed documents and templates aimed at standardizing the way data is entered and structured
- metadata: data about data (i.e., depth, instrument type, location)
- data repository: a place where data is stored and maintained

ESS-DIVE

- ESS-DIVE is a data repository that stores earth and environmental data from DOE-funded projects
- ESS-DIVE developed reporting formats aimed at standardizing data being stored with ESS-DIVE
- Reporting formats are being introduced to WFSFA data providers

Watershed Function SFA (WFSFA)

- WFSFA is a project based in the East River Watershed, located near Crested Butte, CO
- The East River is a diversified testing site where diverse data are collected
- WFSFA's goal is to determine how perturbations affect watersheds
- Various data collection and modeling is conducted as part of this DOE-sponsored project

EAST RIVER WATERSHED

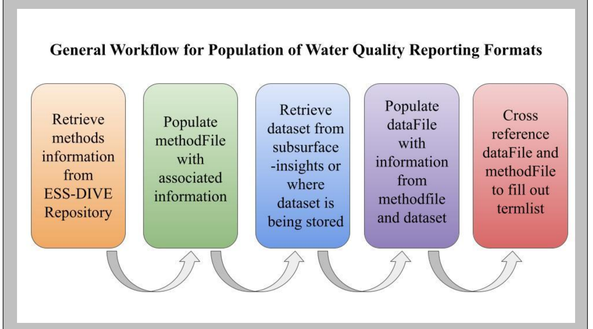
Snowy Peak of East River Watershed

Photo Credit: WFSFA

Image of East River Watershed

Photo Credit: WFSFA

WORKFLOW



- metadata information accessed from ESS-DIVE
- datafiles accessed via ESS-DIVE or Subsurface Insights [a database]
- files stored on google programs

REPORTING FORMATS

#Columns	7						
#Header_Rows	11						
Field_Name	Sample_Name	IGSN	Material	dateTime	deltao18_none	deltad_none	data_notes
Unit	N/A	N/A	N/A	YYYY-MM-DD	ppt_(per-mil, %)	ppt_(per-mil, %)	N/A
Unit_Basis	N/A	N/A	N/A	N/A	parts_per_thousand	parts_per_thousand	N/A
MethodID_Storage	N/A	N/A	N/A	N/A	Stor_sample	Stor_sample	N/A
MethodID_Preservation	N/A	N/A	N/A	N/A	Refrig_sample	Refrig_sample	N/A
MethodID_Preparation	N/A	N/A	N/A	N/A	samplePreparation_04	samplePreparation_04	N/A
MethodID_Analysis	N/A	N/A	N/A	N/A	Anl_sample		N/A
Analysis_DetectionLimit	N/A	N/A	N/A	-9999	-9999	-9999	-9999
Analysis_Precision	N/A	N/A	N/A	-9999	0.025	0.10%	-9999
MethodID_DataProcessing	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Data_Status	N/A	N/A	N/A	raw	processed	processed	N/A
#Start_Data	tuttleSpring_2016-10-19	N/A	stream water	2016-10-19	-18.369955	-136.069271	N/A
N/A	tuttleSpring_2016-10-27	N/A	stream water	2016-10-27	-18.580289	-137.032112	N/A
N/A	tuttleSpring_2016-11-23	N/A	stream water	2016-11-23	-18.260694	-135.751708	N/A

dataset transferred to current water quality reporting format

Method_ID	Method_Type	Method_Description
sampleCollection_01	Sample_Collection_source	Samples are collected directly from the source for the tributary streams and groundwater monitoring wells (following pumping)
sampleCollection_02	Sample_Collection_automatic	Samples are collected from an automated sampler bottle in the case the Pumphouse-ISCO and Coal 11-ISCO surface water location.
sampleCollection_03	Sample_Collection_groundwater	Samples are collected from groundwater monitoring wells through above ground or below ground pumps used to recover fluids from the well
sampleCollection_04	Sample_Collection_isotope	Samples are collected for isotope measurement
precip_collection	Precipitation_Sample_Collection	Precipitation samples were collected at 54 Paradise Road, Mt. Crested Butte, CO, 81225 on an event basis (snow or rain, depending upon season) over the 7-31-2014 to 10-14-2016 period.
Filt_sample	Sample Filtration	Samples are filtered through 0.45-micrometer syringe filters into
Stor_sample	Sample Storage	Stored in sealed vials
Refrig_sample	Sample Storage	Refrigerated until analysis
Anl_sample	Sample Analysis	Sample aliquots (1500-microL) are added to auto-sampler vials fitted with septa-sealed caps and loaded onto an autosampler prior to analysis using either a Picarro L2130-i or Los Gatos Research Liquid Water Isotope Analyzer (LWIA) both of which operate using the same principle of laser absorbance spectroscopy using a cavity ringdown detector. Both instruments provide a precision of 0.025 and 0.1% for delta18O and delta2H, respectively.

metadata accessed on ESS-DIVE entered into methodFile

CONCLUSION

Deliverables

- ❏ Reporting formats examples were developed for WFSFA data providers
- ❏ Feedback was given to reporting format developer
- ❏ Community post and ESS-DIVE webinar in the works

Impacts

- ❏ Reporting format easier to use
- ❏ WFSFA team closer to adopting reporting formats across multiple research projects across East River Watershed
- ❏ WFSFA data structured in more consistent format

Future Work

- ❏ Adapting and applying reporting formats across all ESS-DIVE datasets
- ❏ Future integration of feedback from data providers
- ❏ Integration of reporting formats across all data providers within WFSFA

ACKNOWLEDGMENTS

Special thanks to Rob Crystal-Ornelas, Charuleka Varadharajan, ESS-DIVE team, ESS-DIVE Community Partners, Charu's Data Science team, WFSFA team, and WD&E staff for their support during the term of the internship.

This work was supported in part by the U.S. Department of Energy, OFfice of Science, Office of Workforce Development for Teachers and Scientists (WDTS) under the Community College Internship (CCI) program.