Microorganisms in the Critical Zone: An ASSET Case Study for Sample Processing and Data Workflows, Synthesis, and Dissemination

E.L. ARONSON[1]*, M. MALTZ[1], D. GARIJO[2], K. AROGYASWAMY[1], N. DOVE[3], E. MAYORGA[4], H. FREUND[1], C. BRUYERE[5], M. DANIELS[5]

[1]UC, Riverside, CA, US (*correspondence: emmaa@ucr.edu)
[2]Information Sciences Institute, USC, Los Angeles, CA, US
[3]Oak Ridge National Laboratory, TN, US
[4]University of Washington, Seattle, WA, US
[5]NCAR/UCAR, Boulder, CO, US

Workflows for the documentation, discovery, reconciliation, co-analysis, and dissemination of environmental and microbial data vary greatly between scientists and laboratories. Many sample and data processing workflows are poorly documented and therefore difficult to reproduce. This lack of consistency has led to difficulties in data discovery, analyses, syntheses, and management planning. A case study, the EarthCube Critical Zone Integrative Microbial Ecology Activity (CZIMEA), was initiated in an effort to characterize microbial community structure and function throughout the soil profile at all former US Critical Zone Observatories (CZOs). This project brought together dozens of interdisciplinary scientists to collect soil samples in 10 cm increments down to 1 m or refusal. Collaborators and participants then conducted a variety of analyses on these samples, including, but not limited to amplicon and shotgun metagenomic sequencing, microbial biomass, isotopic and elemental analyses, and potential extracellular enzyme activities. The resulting data and metadata were integrated into NSF-funded EarthCube and related cyberinfrastruture (CI), for subsequent data analyses, visualization and distribution.

The EarthCube ASSET Workflow Sketching Tool was particularly helpful for facilitating communication among interdisciplinary project members and other scientists. It allows for streamlined decision-making among project collaborators and continues to adaptively inform sophisticated dissemination techniques of our project's workflows, beyond the methods sections of our publications. We also explored the use of ASSET for workflows involving physical samples, which can capture and preserve the relationships between related samples and subsequent data generated from those samples in tandem with the current capabilities of International GeoSample Numbers (IGSNs). Beyond ASSET, we discovered that due to disparate repositories and poor compatibility in standards between environmental and microbial data from the same set of samples, there is a pressing need for interdisciplinary tools to connect and translate between repositories and allow for holistic data co-analysis.