Geophysical Research Abstracts
Vol. 21, EGU2019-4496, 2019
EGU General Assembly 2019

Planet Microbe: Toward the integration of oceanographic ‘omics, environmental and physiochemical data layers

Alise Ponsero, Kai Blumberg, Matthew Bomhoff, Ken Youens-Clark, and Bonnie Hurwitz
University of Arizona, Biosystems engineering, Tucson, United States (alise.ponsero@gmail.com)

Oceanographic research cruises provide abundant physical, chemical, and biological data, using a wide range of methods and equipment, very often through the collaborative efforts of scientists from a range of disparate disciplines. These research endeavors span across a broad array of disciplines and are critical to investigating questions related to the interplay between biological, geological, and chemical processes in the ocean systems over space and time. The advent of sequencing technologies allows for an analysis of gene expression in a variety of environmental settings, to measure the distribution and significance of metabolites and lipids in organisms and the environment.
Importantly, although scientists carefully curate and share their data with collaborators to advance individual studies and publications, no systematic, unifying framework currently exists to integrate ‘omics data with physical, geological, geochemical, and biological datasets standardly used by the broader geoscience community. In this context, the development of resources to promote the publication of oceanic ‘omics datasets along with their physio-chemical information is critical.

Here, we present Planet Microbe, a federated resource to enable data discovery and open data sharing for historical and on-going oceanographic sequencing efforts. In this project, several historical oceanographic ‘omics datasets (Hawaii Ocean Time-series (HOT), Bermuda Atlantic Time-series (BATS), Global Ocean Sampling Expedition (GOS), Dark Energy Biosphere Investigations (C-DEBI)) will be integrated into Planet Microbe and reconnected to their physiochemical measurements. New oceanic large-scale datasets such as the Tara Ocean Expedition and Ocean Sampling Day (OSD) will also be integrated into the platform. To systematically integrate these data, we make use of and extend existing ontologies to provide a unique search interface for data discovery based on physiochemical parameters and features in these disparate datasets.

Moreover, the Planet Microbe project aims to provide community tools to support the publication and analysis of new oceanic datasets. Modern data publication requires the user to meet FAIR data standards to ensure data are Findable Accessible, Interoperable, and Reusable (FAIR). However, few tools are currently available for the user to ensure the data are meeting these important standards. Planet Microbe aims to provide tools to evaluate private datasets against these standards and promote the integration of new data in standard data repositories.

Finally, Planet Microbe provides researchers with the infrastructure required to describe and store sequence data, discover and link data sets by important contextual metadata, and save and share project outputs. Members of the research community can integrate tools and pipelines using National Science Foundation (NSF) funded Cyberinfrastructure (CyVerse and XSEDE) to provide users with free access to large-scale computing power to analyze and explore their datasets. These cyberinfrastructure components are broadly useful to any research community.