

OceanMicrobe, a DOME Sample-Data Platform

Jiasong Fang Shanghai Ocean University

DOME will build a planet-scale sample-data platform, *OceanMicrobe*, following the FAIR (Findable, Accessible, Interoperable, and Reusable), and CARE principles (Collective benefit, Authority to control, Responsibility and Ethics), ensuring interoperability with data generated by other Decade programmes and projects.

OceanMicrobe integrates molecular microbiological data with geological, physical, and geochemical datasets, allowing us to study deep ocean microbial diversity and biogeochemical processes. OceanMicrobe will store digital information, including molecular data (metagenome-assembled genomes (MAGs), contigs, genes, etc.) across habitats, geography, and phylogeny with environment data, to facilitate true planetary scale studies of ocean microbial diversity, function, and processes. We aim to address differences in data formats, units, and quality among different datasets, ensuring data consistency and comparability, by employing advanced data integration technologies, including ETL tools, metadata management, semantic integration, service-oriented architecture, etc. Additionally, the platform will use cloud computing, information management, data standards and protocols, and data mapping architecture to enable data visualization, interaction, customization, and scenario analysis. OceanMicrobe will interoperate with other resource databases such as proGenomes, OBIS, GOOS, GMGC, metaMap, etc. and provide equal open access to samples, specimens, data, knowledge, technology and skills to everyone. The generated data will also provide a fundamental basis for conservation and sustainable use of ocean genetic resources, supporting the implementation of BBNJ Agreement.

During the execution of the mega-project (Developing world-leading technologies for exploration and exploitation of hadal biological resources and conducting frontier research on biological processes and genetic resources, funded by Ministry of Science and Technology of China, total budget CNY 18.53 million, 08-2018–12-2021), directed by Dr. Fang (Chief Scientist), we built a data platform called HADAL (see Figure 1 below). This platform hosts biological, genetic, and environmental data, and also has bioinformatic analysis capabilities, including analysis of 16S rRNA data and metagenomics data, and functional annotation. This platform will be used as a model for building the *OceanMicrobe* platform.





Figure 1. The HADAL data platform.