



## OSD Data Policy

The novel Ocean Sampling Day (OSD) concept was designed within the EU-funded project Micro B3. In 2018, EMBRC ([www.embrc.eu](http://www.embrc.eu)) decided to offer long-term support to the continuation of OSD. Through the ASSEMBLE PLUS project, EMBRC will support the transition of the logistics and operation of OSD. Being a self-sustained European Infrastructure, EMBRC can secure the long-term operation of OSD and its step-by-step expansion to additional ecosystem components and in coordination with the global Genomics Observatories consortium.

The Institute of Marine Biology, Biotechnology and Aquaculture of the Hellenic Centre for Marine Research in Heraklion, Crete (IMBBC-HCMR), will coordinate the upcoming event of 2019 and will be the recipient of samples to be submitted for sequencing.

The OSD Data Policy was authored by the OSD Data Policy Working Group within MicroB3-EU project and it was finalized at the Micro B3 General Assembly (GA) April 2014. The present version has been adapted from “MicroB3 Ocean Sampling Day Handbook, version of April 2016” and has been modified accordingly for the Ocean Sampling Day events, that will be coordinated by the Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research.

### Background

The OSD Data Policy was created by drawing on existing policies (See Biosharing: <http://biosharing.org/policies>) and current best practice in genomics and the needs of Micro B3 and the OSD community. It was developed and will be maintained according to the “12 steps to a data policy” outlined in *Field et al (2009) Science*:

### **Tenets of Ocean Sampling Day (The policy applies to all parties in the OSD Consortium. It is a prerequisite for taking part in OSD events)**

The OSD data set is generated by the members of the OSD Consortium - to be part of the Consortium you must be registered and substantively contribute to the collaboration.

The OSD dataset will be a reference data set and should be as widely accessible and used to support downstream research as possible.

As OSD is one event it should be easier to get all contributors (their official sites) to release the data publicly.

All OSD data should include metadata, at least the mandatory mentioned in the OSD Protocols.

All OSD genomic data must be tightly linked to environmental data - contextual information is a top priority.

Sequence data without deposition of accompanying metadata (minimum of lat, long, time, depth, temp, salinity according to OSD Protocols) will not be released after the June solstice 2020 event. The OSD Consortium will work towards release of all data.

The present OSD data policy covers all future OSD events including the June solstice 2020.

Data will be submitted to public repositories (e.g. INSDC).

All data will be processed/stored in the EMBRC-ERIC e-infrastructure as it is built and the full data set will eventually be available through that infrastructure. Data can flow in many analysis pipelines/databases - we welcome its wider use.

All samples must be collected legally (with appropriate permits e.g. ABS) and individual collectors are responsible for making sure of this.

## **The OSD Consortium**

This data policy applies to everyone in the OSD Consortium (to date this means everyone who has constructively contributed to OSD sampling events).

## **Release of the data and metadata**

We will follow the spirit of previous Ocean Sampling Day Events by supporting Data open sharing. We will work towards: "immediate release of all OSD data, as soon as sequenced and quality checked, to the public". In this sense, "public" means beyond AssemblePlus, EMBRC and OSD participants, to the world at large.

To protect the rights of the consortium to publish the first global analysis of the data, the data will be released under [Ft Lauderdale principles](#). This is routinely used in large sequencing projects and promotes the use of the data by the wider community, while safeguarding the scientific contributions of the data generators. For example, a complete Neanderthal genome

was released to the public prior to publication in 2013 under these principles (see: <http://www.eva.mpg.de/neandertal/index.html>).

This statement will accompany any access to all OSD sequence data.

## Use of the sequence data

All data will be made freely available. However, we ask users to observe the [Ft. Lauderdale principles](#), which entitle the data producers to make the first presentation and publish the first genome-wide analysis of the data. The data can be used freely for studies of individual genes or other individual features of these sequences.

## Collaborative Data Analysis

We hope that the OSD reference data set will be helpful in a variety of ways and we encourage those interested in analysis of the data to declare their interests and formally join the OSD Data Analysis working group. Together, we are working towards “OSD reports” that will be generated following each event. Ideally, each report will provide a high-level overview of the event and, if multiple events are available, comparison and contextualization with past data. Predictions about future events may also be included. It is our vision that this type of transparent collaboration where interested parties declare how they would like to contribute to the data analysis, will help maximize the efforts of the community and build the strongest possible biological interpretation of the data.

## Publication

We expect the authorship of the global analysis of the OSD data set to belong to the OSD Consortium. Authorship will be defined on a case-by-case basis for more specialist publications.

## Signed by

Site ID: .....

Site Name: .....

Institution: .....

Country: .....

Site Coordinator: .....

Signature: .....

Date: .....