Marine

Microbial Biodiversity,

Bioinformatics & Biotechnology

Grant agreement n°287589

Acronym : Micro B3
Start date of project: 01/01/2012, funded for 48 month

Deliverable 2.3

Report of the possible content of the OSD call including a first version of the Micro B3 data policy

Version: 1.0
Circulated to: WP2, All WP Leaders (June 2013)
Approved by: Frank-Oliver Glöckner (June 2013)

Expected Submission Date: 31.03.2013
Actual submission Date: 01.07.2013
Lead Party for Deliverable: UOXF

Mail: osd-contact@microb3.eu

Dissemination level:	
Public (PU)	x
Restricted to other programme participants (including the Commission Services) (PP)	
Restricted to a group specified by the consortium (including the Commission Services) (RE)	
Confidential, only for members of the consortium (including the Commission Services) (CO)	

Summary

In this deliverable we give an update on Micro B3 plans to issue on OSD call and provide a first version of the OSD Micro B3 data policy. Both were discussed among WP Leaders at the EEB meeting in Bremen May 6-8, 2013. In summary, we decided 1) due to the successful recruitment of sites in the OSD pilots to replace the call with a targeted spend to support OSD 2014, 2) to form an OSD steering committee to finally decide on the spending of the money and 3) to base our OSD data policy on open and public release of all OSD data as soon as it is sequenced. All three items are open for consultation and further refinements up to the next General Assembly planned for April 2014.

I. Call for participation in OSD

When we originally envisioned the Micro B3 project we placed the Ocean Sampling Day (OSD) event at month 30. We have now run pilot OSD events (as described in previous deliverables 2.1 and 2.2) and with the strong engagement of the community are rethinking our OSD strategy. In the proposal we planned to write a call for participation and back the call with discretionary funding for sites to undertake samples. We planned to use a variety of criteria to judge sites and release funds according to criteria including 1) the amount and quantity of existing molecular data, 2) the link to environmental and other contextual data 3) the novelty of the site 4) the strategic value of the site with respect to creating a comprehensive survey of sites across Europe (and beyond). To evaluate our choice of sites for inclusion in OSD, we proposed to form an expert Selection Committee, and the funded sites would be the OSD 2014 event (additional sites could also join in on an unfunded basis).

Due to the successful recruitment of 60 sites (for June 2013) we have decided at the Micro B3 EEB meeting in Bremen May 6-8, 2013 that we would replace the action of having a call with the creation of an expert committee to decide how to allocated the funds against data generation, especially sequencing. This is now possible because of the tremendous in kind contributions of our sampling sites.

We now propose to go forward as follows:

1) Form an Micro B3 OSD Steering Committee with the leaders of the Micro B3 WPs
2) The OSD Core Team (WP2) puts together a proposal for spending the discretionary funding on the 2014 event
3) Micro B3 OSD Steering Committee to vote on action to spend the budget at the next Micro B3 General Assembly planned for April 2014
4) Spend the OSD budget to support OSD 2014 event.

In considering the budget, we also have decided to pull forward at least a small amount of funds to prioritize metagenomic (shotgun) sequencing of a number of metagenomes. We are currently discussing this with Argonne National Labs. We would aim to sequencing 20 metagenomes from the sites sampled to date based on our criteria above (i.e. best contextual information, but also good DNA yields, etc). We expect the cost of this sequencing effort to be less than $10k. This metagenome sequencing will generate strategic data for validating the existing and planned Micro B3 informatics pipeline.

II. Towards and Micro B3/OSD data policy

It is essential that the OSD Consortium understand how the data will be shared/used. We are therefore creating a data policy that defines the responsibilities of all involved. We are developing this policy in accordance with the methodology outlined in Field et al (2009) (see Annex I). *Science*. According to this process we have been working to define the scope of the policy over a series of meetings.

We discussed the initial data policy at the May 2013 EEB meeting and further formalized our path.
We agreed that:

1) By design we will aim to create an OSD data policy that is as simple as possible. Our default in all cases will be open publication and access to data. We will continue to work very closely with WP8 (IPR) on the creation of an OSD data policy.

2) To extend the initial policy document and transform it into the first version of a formal policy. Our revised draft now includes our agreements on 1) releasing all data immediately and to the public as soon as sequenced (we will consider a Creative Commons license) 2) forming an OSD Consortium and 3) continuous updates to better reflect details of the workflow of OSD within Micro B3 and with engaged partners 4) addition of sections on plans for analysis and publication.

3) Formalize a data policy working group (WP leads of Micro B3)

4) Put the current draft policy out for consultation with the OSD Community until three months from the Micro B3 General Assembly in April 2014

5) Ratify the OSD Data Policy at the next Micro B3 General Assembly meeting

We have initiated all actions and will prepare the final policy document by January 2014. The current draft of the OSD Data Policy along with the members of the OSD Data Policy Working group are now available (see Annex II) and details have been circulated to the OSD community.

ANNEX 1

Draft Micro B3/OSD Data Policy
(Last updated June 2013)

Authored by the OSD Data Policy Working Group with the OSD Consortium

OSD Data Policy Working Group

Dawn Field, Oxford University, OSD, WP2 Leader
Mesude Bicak, Oxford University, OSD Co-ordinator/Bioinformatician
Anna Klindworth, MPI Bremen, OSD Co-ordinator
Frank Oliver Glöckner, MPI Bremen, Micro B3 Co-ordinator
Arianna Broggiato, BIOGOV, ABS (WP8)
Arul Scaria, BIOGOV, Data Policy (WP8)
Tom Dedeurwaerdere Univ Louvain WP8 Leader
Alban Ramette, AWI, Biostatistics (WP5 & WP6)
Linda-Amaral Zettler, Woods Hole, MicroB3 Scientific Advisory Role (ICOMM MIRADA LTER)
Petra ten Hoopen, EBI, Metadata & Standards (WP2 & WP4)
Michelle Barbier, CIESM, Med Co-ordinator (WP2 & WP9)
Neil Davies, UC Berkeley, GOs Network, WAML, Pacific Region
Jack Gilbert, Argonne National Labs, EMP
Katie Barker, Smithsonian Institution, GGI Bio-Archiving

Status to date

This draft data policy covering the collection, dissemination, analysis and publication of OSD data has been developed over the past year during the pilot events of OSD. Within Micro B3, WP8 is responsible for the Micro B3 data policy. The OSD policy will be based on this policy. We welcome input from all OSD participants on shaping of the OSD data policy. Please send comments to osd-contact@microb3.eu.

Next steps

This draft data policy is undergoing OSD community consultation until the end of 2013 and will be finalized and voted upon in the Micro B3 General Assembly (GA) planned for April 2014 and then adopted. It will remain the responsibility of the OSD Data Policy working group to make sure the policy is fit for purpose and all are aware of it when submitting and using data of the OSD Consortium. If you would like to join the group please contact osd-contact@microb3.eu.

Background:

We are developing the OSD Data Policy by drawing on existing policies (See Biosharing: http://biosharing.org/policies), current best practice in genomics and the needs of Micro B3 and the OSD community. We are developing the policy according to the "12 steps to a data policy" outlined in *Field et al (2009) Science:*

12 Step Path to a Data Policy:
http://blog.biosharing.org/2010/02/12-steps-to-creating-data-sharing.html

Below we describe the tenets of data sharing in OSD, release of the data, and the membership in and authorship with the OSD Consortium. **The policy will apply to those in the OSD Consortium.**

Tenets of Ocean Sampling Day

The OSD data set is generated by the members of the OSD Consortium - to be part of the Consortium you must be formally 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 (a one-off event) it should be easier to get all contributors (their official sites) to release the data publicly.

All OSD genomic data must be standards compliant - must include metadata - this includes compliance with the GSC's MixS standard.

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

All members of the OSD Consortium are expected to support the OSD data policy.

OSD will ramp up to the official OSD summer of 2014, sampling each solstice (starting summer 2012) - we are already looking into ways to sustain sampling beyond 2014

Data will be submitted to the INSDC (ENA/MG-Portal). Ideally an INSDC bioproject page will be created to keep the data set as a set (collective sequences) (Ilene at NBCI suggested this for OSD)

All data will be processed/stored in the Micro B3 infrastructure as it is built and the full data set will eventually be available through this infrastructure. Data can flow in many analysis pipelines/databases - we welcome its wider use

All samples must be collected legally (with appropriate permits) and individual collectors are responsible for making sure of this. Specifically ABS/MTA/DTA agreements will be in place as required (bioarchiving at GGI requires OSD MTA, proof of relevant ABS and metadata)

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

At the 2013 EEB meeting of the Micro B3 project we agree on the importance of the OSD data set as a whole (the sum of participating sites) and the desire to support open sharing of the data. The EEB agreed to work towards:"immediate release of all OSD data as soon as sequenced to the public". In this, sense "public"means beyond Micro B3 and OSD participants to the public 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 principle (see: http://www.eva.mpg.de/neandertal/index.html).

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

Use of the sequence data

All data is made freely available. However, we ask users to observe the Ft. Lauderdale principles, which entitles 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.

Evolution of this Policy

According to the "12 steps for policy development" process, this policy may evolve according to best practice. As agreed at the May 2013 EEB meeting of Micro B3, changes to the policy will be done in consultation with the community. Specifically, policy issues will be taken forward through the OSD Data Policy Working group who be responsible for building consensus on any issues or changes. Final authority for any changes and for upholding the policy will rest with the leadership of Micro B3.

ANNEX II

Towards an OSD data policy

Within Micro B3, WP8 is responsible for the Micro B3 data policy. The OSD policy will be based on this policy. We welcome input from all OSD participants on shaping of the OSD data policy.

Background:

We will be drawing on existing policies to formulate the OSD policy.

Data policies of major funders supporting 'omic research
http://biosharing.org/policies
12 steps to a data policy
http://blog.biosharing.org/2010/02/12-steps-to-creating-data-sharing.html
concept of a minimum data policy
http://blog.biosharing.org/2010/04/minimum-data-sharing-policy.html

Tenets of Ocean Sampling Day

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

The OSD dataset will be the product of contributions of many independent sites joining the OSD collaboration - joining the collaboration means adhering the OSD data policy and terms of collaboration

OSD will ramp up over the next two years to the official OSD summer of 2014, sampling each solstice (starting summer 2012) - we are already looking into ways to sustain sampling beyond 2014

The default stance is that all data should be released immediately (no moratorium) to group

Data will be submitted to the INSDC (ENA/MG-Portal)

Ideally an INSDC bioproject page will be created to keep the data set as a set (collective sequences) (Ilene at NBCI suggested this for OSD)

All data will be processed/stored in the Micro B3 infrastructure as it is built and the full data set will eventually be available through this infrastructure

We expect to work under a model of group authorship on overview analysis of the OSD data

Specific groups/authors may do specialized analysis of the data as agreed within the consortium

Once the full data set is published it can be used by all for any reasons (collaboration obviously encouraged)

Data will be publicly available and shared

Data will be compliant with the GSC's MixS specification and be as richly contextualized as possible

All contributors have the rights to analyze and published on their own data (local story), the full OSD data set belongs to all OSD contributors

We will make every effort to make the OSD data set as uniform as possible to aid comparative analysis

Data will flow in many analysis pipelines/databases - we welcome its wider use

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

ABS/MTA agreements will be in place as required (bioarchiving at GGI will require this)

All sequencing partners will have access to the data, but will return all sequencing data to the OSD consortium

Release of the data

All data from OSD will be made public

Evolution of this Policy

According to the 12 steps to policy development, this policy may evolve according to best practice. As agreed at the May 2003 EEB meeting of Micro B3, changes to the policy will be done in consultation with the community. Specifically, policy issues will be The OSD Data Policy Working group will be responsible for building consensus on such changes. Final authority for any changes and for upholding the policy will rest with the leadership of Micro B3.

Logic

If considered as one big event (a one-off event) will be easier to get all contributors (their official sites) to release the data publicly

All genomic data must be standards compliant - must include metadata

All genomic data must be tightly linked to environmental data