



# Marine Microbial Biodiversity, Bioinformatics & Biotechnology



Grant agreement n°287589

Acronym : Micro B3

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

## Deliverable 3.4

# Selection and definition of geographical pilot areas for Micro B3 cases

Version: 1.3

Circulated to: WP3 partners, Dawn Field (WP2), Mesude Bicak (WP2), Chris Bowler (WP6), Josep Gasol (WP6) (24.01.2013)

Approved by: Dawn Field and Stéphane Pesant (12.02.2013), Dick Schaap (14.02.2013)

Expected Submission Date: 31.12.2012

Actual submission Date: 14.02.2013

Lead Party for Deliverable: IFREMER

Mail: [catherine.borremans@ifremer.fr](mailto:catherine.borremans@ifremer.fr)

Tel.: +33 (0)2 98 22 41 90

---

### 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

The Micro B3 project aims for a better understanding of the complexity of marine microbial communities and their role in climate change. This requires that the data sets and information on marine organisms and genes are complemented with their environmental context. WP5 is charged with building the Micro B3 Information System (MB3-IS) to provide the bioinformatics capacity for marine biodiversity data processing, analysis and biotechnological exploitation. The MB3-IS analyses are intended to be based on a number of Case Studies (scientific questions we wish to answer using the Micro B3 data integration pipeline) representative of current activities related to exploration of marine microbial ecosystems (and which have specific needs in terms of bioinformatics developments). This requires data input from both the genomic data infrastructure (EMBL-EBI) and the ocean environmental data infrastructure. Oceanographic and marine environmental data will be provided to Micro B3 through the overarching infrastructures, SeaDataNet and EurOBIS, that also are well involved in the EMODNet development. These oversee and give access to extensive volumes and types of data sets from existing ocean and marine data collection activities from multiple sources. Moreover data will be collected in the framework of Micro B3 via the **Ocean Sampling Day (OSD)** and derived from the **Tara Oceans** expedition and **Malaspina cruise**. These data will enable to test the bioinformatic and environmental MB3-IS in practice. The portal and the services which will gather and deliver environmental data in the MB3-IS structure (i.e the interfacing between SeaDataNet-EMODNet CDI service and the MB3-IS) are planned to be shown to work for specific geographical sites. This Deliverable defines those locations, which were selected as Micro B3 genomic and oceanographic data matching areas relevant for the Micro B3 Use Cases.



## Table of Contents

- 1.0 Context of the Deliverable**
  - 2.0 Approach**
  - 3.0 Micro B3 Use Cases**
  - 4.0 Overview of Micro B3 sampling sites provided by participants**
    - 4.1 Ocean Sampling Day**
    - 4.2 Long term sampling sites**
    - 4.3 Tara Oceans expedition (spatial monitoring)**
    - 4.4 Malaspina expedition (spatial monitoring)**
  - 5.0 Environmental data management systems data coverage**
    - 5.1 SeaDataNet-EMODNet**
    - 5.2 EurOBIS and the WoRMS**
    - 5.3 ICES**
    - 5.4 PANGAEA**
  - 6.0 Selection of pilot geographical and temporal sites/areas**
- Reference list**



## 1.0 Context of the deliverable

One of the major challenges in Micro B3 is to organize cooperation and interoperability between the oceanographic research community and the genomics research community. Particularly, WP3 aims at establishing interoperability between the Micro B3 Information System and the Oceanographic Environmental data management systems (see the Deliverable 3.1 for further details).

As a model, relevant services from the environmental data infrastructures will be bundled via a portal, from which the bioinformatics analysis system will be able to harvest data and information in an automated way. The dedicated portal and service will be demonstrated for gathering and making available data sets and data products for a number of geographical pilot areas suitable for the Micro B3 Use Cases.

Those specific sites are defined in the present deliverable and were determined together with WP2 and WP6.



## 2.0 Approach

The pilot locations were chosen to be of relevance for datasets resulting from research activities that are planned to take place in Micro B3, namely the samples analyses and studies from the Ocean Sampling Day-OSD- (WP2), from the long-term monitoring and study sites, and from the Tara Oceans and Malaspina past cruises (WP6).

The data coverages of these data sources and of the oceanographic environmental data management systems (SeaDataNet-EMODNet, including ICES and PANGAEA, and EurOBIS) were crossed to select the most suitable sites for the portal and services working demonstration in the Micro B3 Use Cases context.

The actual gathering and delivery of data sets through the oceanographic services for Micro B3 cases will be reported in the Deliverable 3.6 (M30).

### 3.0 The Micro B3 Use Cases

WP4 will deliver standards for data acquisition and handling that will support interoperability between ocean sampling processes and effective sharing of marine microbial data derived from the sampling. Those concepts development will be fed into by several streams of information, of which the prototype Use Cases are relevant for this Deliverable. The Micro B3 Use Cases are scientific questions/hypotheses in biological or environmental sciences that will be asked of the sampled marine microbial systems. Four biological and two environmental prototype Use Cases were identified by WP4 and the assistance of the Micro B3 Consortium partners. The biological Use Cases focus on diatom biology and the environmental Use Cases on the marine prokaryotic biodiversity (see the Deliverable 4.1 and The Use Case Document for further details and a summary of the prototype Use Cases).

The Use Cases were transformed into a set of scientific and legal parameters -called collectively the Micro B3 Candidate Checklist- that are needed to answer the questions postulated in the Use Cases.



## 4.0 Overview of Micro B3 sampling sites provided by participants

Appropriate samples are crucial for the research activities to be conducted in Micro B3. The following samples providers will bring a rich set of sequence and environmental data, on both temporal and spatial scales, for in depth bioinformatics analysis in the Use Cases context.

### 4.1 Ocean Sampling Day

The Ocean Sampling Day (OSD) is a massive sampling event planned on summer solstice (June 21<sup>st</sup>) in the year 2014. This project will provide insights into fundamental rules describing microbial diversity and function. Indeed, these cumulative samples, related in time, space and environmental parameters, will contribute to determine a baseline of marine biodiversity and functions on the molecular level. In preparation of OSD, WP2 -which coordinates this project- is organizing sampling at each solstice until the main event in 2014. Best practices (uniform sampling protocols) and a list of suitable sampling sites are being produced in order to ensure maximum usefulness of the samples and chances of success. All of the information in the sampling sites registry and any sequence data generated from the OSD samples will be put into the public domain. This will be through the Micro B3 Catalogue and the Micro B3 Information System (WP5).

The initial list of suitable sampling sites is being built and will be reported in the Deliverable 2.2. This is the sites list established on the basis of the information currently available:

Site	Location	Latitude	Longitude
L4	English Channel (UK)	50.25	-4.22
Roscoff	English Channel (France)	48.78	-3.94
Helgoland	North Sea (Germany)	54.19	7.90
Naples	Naples (Italy)	40.81	14.25
Heraklion-M3A	Cretan Sea (Greece)	35.78	24.92
Blanes	Spain	41.67	2.80
Moorea	French Polynesia	-17.50	-149.83
BATS	Bermuda	31.61	-64.28
SPOTS	LA (USA)	33.58	-118.38
Rothera	Antarctic	-67.57	-68.13
Churchill	Canada (Arctic)	58.73	-94.07
Banyuls	Med Sea (France)	42.49	3.14
Villefranche sur Mer	Med Sea (France)	43.68	7.32
North Cyprus-Girne	Cyprus	35.36	33.29
Matis	Faxafloi (Iceland)	64.08	-22.13

### 4.2 Long term sampling sites

Temporal monitoring programmes include L4 (Plymouth), Blanes Bay (Barcelona), Naples, Heraklion, Iceland (MATIS) and Mediterranean deep-sea hypersaline anoxic lakes (DHALs) long term sampling sites. Those sites offer good conditions for plankton biology and ecology studies to which genomics studies can be related. They were regularly sampled for the past 10 (or even more) years and the resulting samples are available at the partners institutes. For details about the sites please refer to Annex 1 of the Micro B3 DoW.



### 4.3 Tara Oceans expedition (spatial monitoring)

The Tara Oceans project was launched in September 2009 for a 3-year study of the global ocean ecosystem aboard the ship Tara. A unique sampling programme encompassing optical and genomic methods to describe viruses, bacteria, archaea, protists and metazoans in their physico-chemical environment has been implemented. The goal is to generate open access datasets to be used in probing the morphological and molecular makeup, diversity, evolution, ecology and global impacts of plankton on the Earth system. The sampling strategy, sample/data analysis and data management have been carefully tailored and integrated toward this overarching goal.

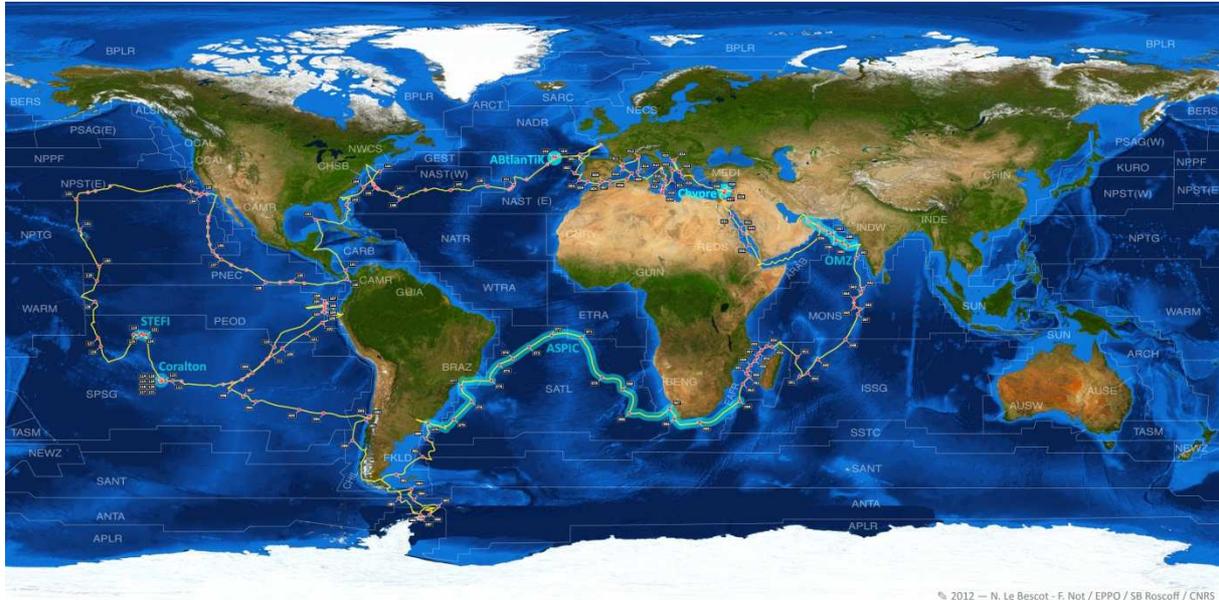


Figure 1. TaraOceans cruise track indicating the sampling stations (154 in total) and highlighting the topical experiments (light blue) carried out during the expedition. A representative of Longhurst's biogeographical provinces has been represented as a background (map credits to Noan Le bescot & Fabrice Not/EPPO/SB Roscoff/CNRS).

The different groups of organisms targeted by Tara Oceans were separated based on their size, using various meshes on the sampling instruments or on filtration units onboard (Figure 2). Tara Oceans has given a particular focus on planktonic protists, including samples for total DNA and RNA, various ways of preserving the communities for laboratory morphological or morpho-genetic analyses (HTM, SEM, TEM, OM, FISH, SAGs), and onboard high-throughput automated imaging (FlowCam, FlowCytometry). Importantly, data and samples were collected from 4 independent size-fractions covering the entire range of protistan biodiversity.

Details of the sampling procedures and sample treatments onboard will be published in a Methods paper (Not *et al.*, in prep).

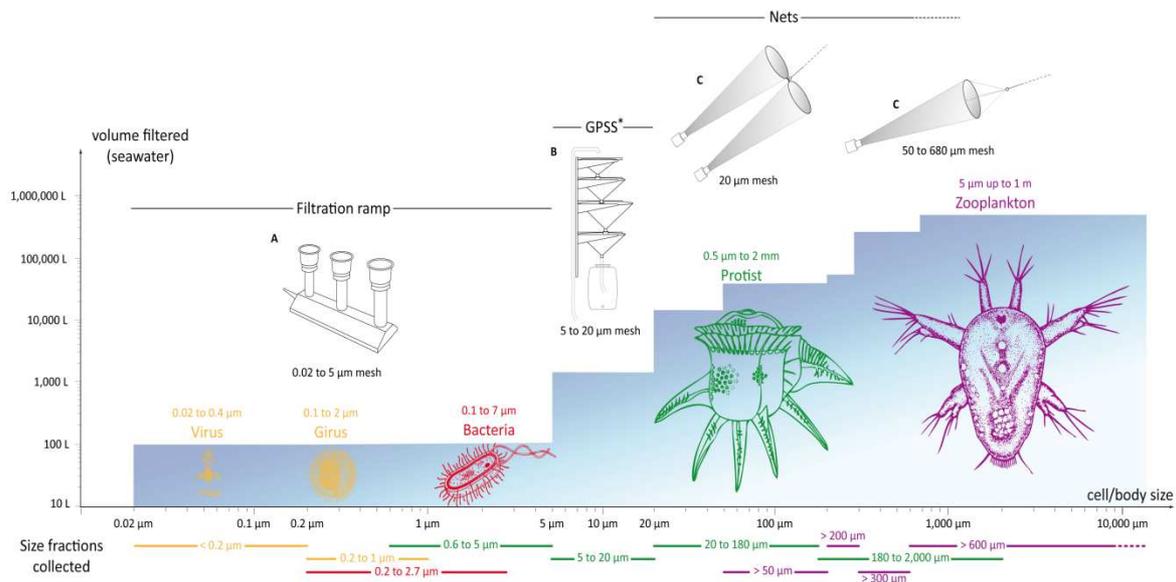


Figure 2. A partial summary of sampling instruments and targeted organisms (Karsenti *et al.* 2011).

Environmental conditions at each sampling location were determined onboard using sensors deployed on the meteorology station (i.e. continuous time series), rosette system (i.e. vertical profiles from 0-1000 m), large volume pumping system (i.e. time series at discrete sampling depths), underway system (i.e. continuous time series at 5 m), and occasionally on surface drifters and gliders. Additionally, the oceanographic context at each station is characterized using remote sensing products such as ocean colour, SST and SSH to determine meso- to large-scale features and variability in key parameters such as surface temperature, salinity, chlorophyll a, currents and mixed layer depth.

### Sampling Strategy

Each station was planned a few days or a week in advance by a team of physical and biological oceanographers, together with the chief scientist on board. This short term planning was chosen to take advantage of the latest oceanographic information available based on processed and analyzed satellite data (Chl, SST and altimetry). Near real-time updates of the satellite images were sent to the chief scientist on board. Furthermore, continuous surface measurements (Temperature, Salinity, Fluorescence) were used to fine tune the sampling locations across fronts or filaments for example. When needed, a preliminary CTD transect was performed to characterize station at meso-scale. Finally, on board analysis of sensor readings from the rosette (e.g. CTD, Oxygen, Nutrients, UVP) was used to identify and target features of special interest in water column, such as DCMs, Oxygen Minimum layers, mesopelagic features, etc. In addition to the general sampling strategy outlined above, some topical studies addressed specific scientific questions and required additional sampling approaches and instruments (i. e. state of the art oceanographic instruments such as gliders, biogeochemical autonomous floats, ARGO floats with drogues and LADCPs) were deployed to improve the success of the survey of the oceanographic feature.

This study of the world oceans microorganisms biodiversity combined classical analysis methods and genomics and is then particularly relevant for the Micro B3 project.

The environmental data and the registry of samples collected during the Tara Oceans expedition are archived and managed centrally at PANGAEA. As the data management for the Tara Oceans cruise is planned to be supported by the Micro B3 project (see Deliverable

3.2), oceanographic services from PANGAEA will be connected to the overarching oceanographic data infrastructures in Micro B3 (SeaDataNet, EurOBIS). For the Tara Oceans expedition, the PANGAEA samples registry will be the key to link samples, data archived in a distributed network of databases and metadata about sampling and analysis methodology.

#### 4.4 Malaspina expedition (spatial monitoring)

Like the Tara expedition, the Malaspina expedition (2010-2011) produced worldwide samples from marine microbial ecosystems, explored on both temporal and spatial scales. Those biological and (meta)genomic samples, correlated with their oceanographic environmental context, will be exploited in Micro B3 to develop an understanding of the ecosystems biology by relating biodiversity with the functional structure of the ecosystem.

The Malaspina global expedition cruise started on Dec 14<sup>th</sup> 2010, and was completed on July 15<sup>th</sup>, 2011. The cruise was divided into 7 legs (Cádiz-Rio de Janeiro, Río-Cape Town, Cape Town-Perth, Perth-Sidney, Auckland-Honolulu, Honolulu-Cartagena de Indias, and Cartagena de Indias-Cartagena), crossing the North Atlantic (Legs 1 and 7), the South Atlantic (Leg 2), the South Indian Ocean (Leg 3), and the South (leg 5) and North Pacific (leg 6). The cruise also sampled Southern ocean waters, south of Australia (Leg 4).

The cruise research activities were divided into several blocks, among them the “Microbial activity and diversity” block is relevant for Micro B3.

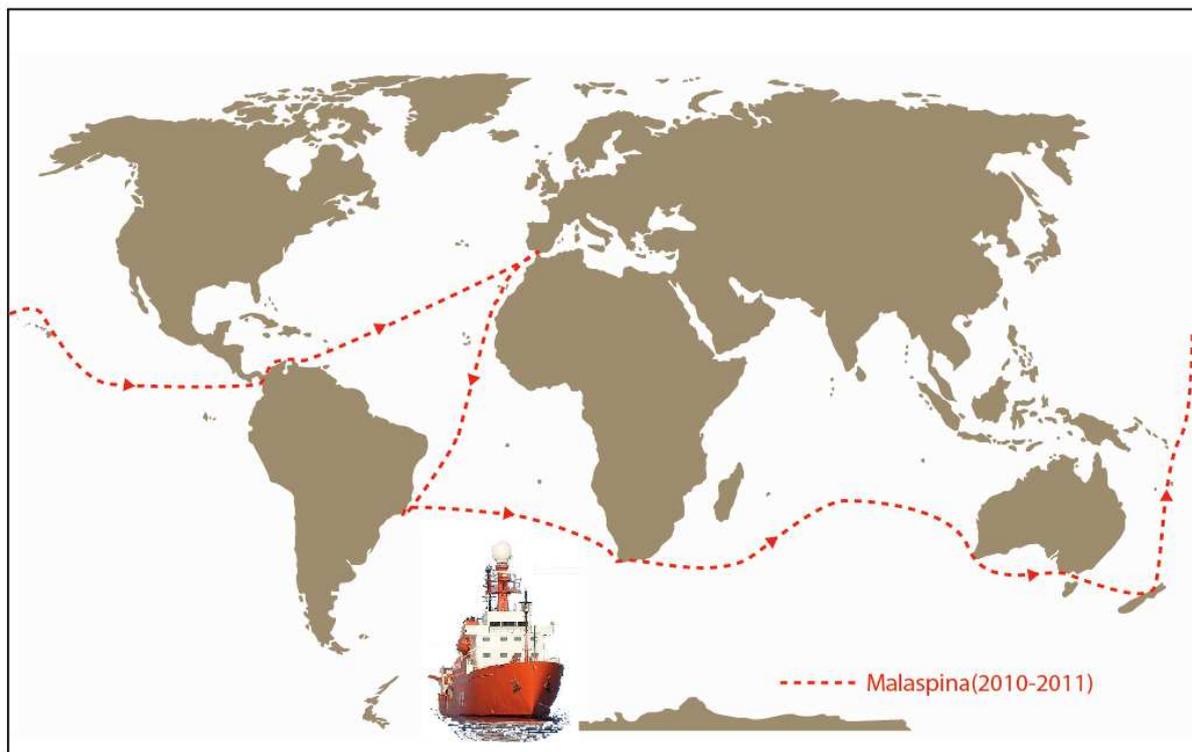


Figure 3. Malaspina cruise track.

Sampling was structured with 1 station a day, with 2-3 CTDs. The stations were classified as “Full profile” or as “Dedicated”. The “Dedicated” stations were divided into three types:

- stations dedicated to MICROBIOLOGISTS (metagenomes, metatranscriptomes, metaproteomes),
- stations dedicated to BIOGEOCHEMISTRY (Organic matter fingerprinting),
- stations dedicated to both.

The following is a list of all the variables analyzed by the Microbial Block (not all were taken from all samples):

- Bacterial and picoalgal abundance (by flow cytometry)
- Bacterial physiological status (NADS, CTC)
- Bacterial size (sequential filtration and flow cytometry)
- Virus abundances (flow cytometry)
- Viral morphological diversity (Transmission electron microscopy)
- Protist abundances (flow cytometry)
- Protist abundances (DAPI-epifluorescence)
- Deep Bacterial Respiration ( $O_2$  consumption)
- Bacterial activity/production (surface: leucine incorporation)
- Bacterial/Archaeal activity/production (deep samples: leucine incorporation and use of archaeal inhibitors)
- DNA sampling with 0.22  $\mu\text{m}$ , 0.8  $\mu\text{m}$  and 3  $\mu\text{m}$  filter sizes
- RNA sampling with 0.22  $\mu\text{m}$ , 0.8  $\mu\text{m}$  and 3  $\mu\text{m}$  filter sizes
- Viruses sampling by precipitation of 0.22  $\mu\text{m}$  filtrate (Sullivan's method)
- Bromo-diuridine-incubated samples. Collection of DNA for BUMP-analyses
- Metaproteome samples
- Samples for SAGs (Single-cell amplified genomes)
- Exoenzymatic activities
- ECO-Biologs for bacterial metabolic diversity profiles
- Transparent exopolymeric particles (TEPS)
- FISH samples for prokaryotes
- FISH samples for eukaryotes
- Samples for Bacteriochlorophyll  $a$  determination
- Samples for AAP determination (special filters + DAPI)

Additionally, several sets of experiments were performed at selected stations. These were:

- Grazing by protists
- Viral mortality
- TEP formation and degradation
- Size distribution of organisms and particles
- Inorganic C incorporation by prokaryotes
- $^{14}\text{CO}_2$  and  $^3\text{H}$ -leucine incorporation by archaea and bacteria
- Nutrient limitation experiments
- Heterotrophic light-use experiments

Concerning Malaspina data, they are firstly stored in Malaspina Digital, a dedicated site ([www://metamalaspina.imedeo.uib-csic.es/geonetwork/srv/es/main.home](http://www://metamalaspina.imedeo.uib-csic.es/geonetwork/srv/es/main.home)). Then they will be sent either to PANGAEA or to the IEO Spanish Data Center (decision to be taken in spring 2013) and will that way become available in the Micro B3 project context.



## 5.0 Environmental data management systems data coverage

For a detailed description of each data management system cited here under, see Deliverable 3.1.

### 5.1 SeaDataNet-EMODNet

SeaDataNet has developed an efficient distributed Marine Data Management Infrastructure for the management of large and diverse sets of data deriving from in situ and remote observation of the seas and oceans. Professional data centres, active in data collection, constitute a Pan-European network providing on-line integrated databases of standardized quality. SeaDataNet infrastructure and standards were adopted as basis for the implementation of the EMODNet project. This system will ensure interoperability and harmonization between the six EMODNet lots (hydrography, chemistry, physics, biology, geology, habitats).

In Micro B3, SeaDataNet will be a major source of ocean and marine environmental data to complement datasets and information on organisms and genes.

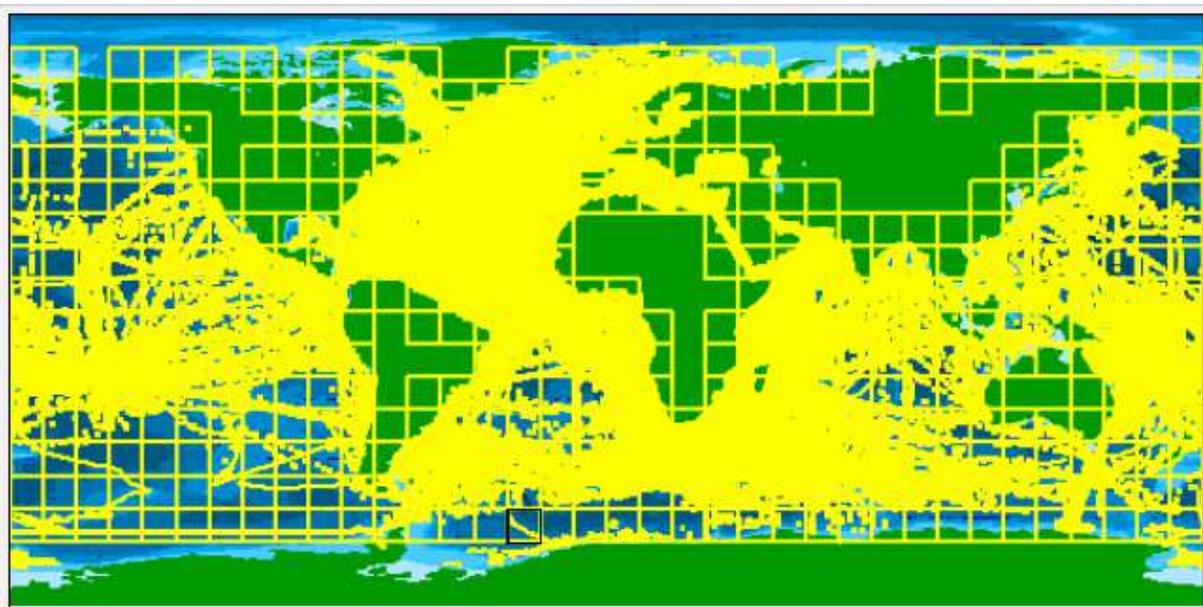


Figure 4. SeaDataNet data coverage (points, tracks and areas) for the timescale 1950-2012.

### 5.2 EurOBIS and the WoRMS

EurOBIS acts as the European node of OBIS. It is a distributed information system giving access to biogeographic data on marine species collected by European institutions. EurOBIS refers to the WoRMS (the World Register of Marine Species) taxonomy, which will also be the standard for Micro B3 species data.

EurOBIS will provide the biodiversity information corresponding to Micro B3 specific data.



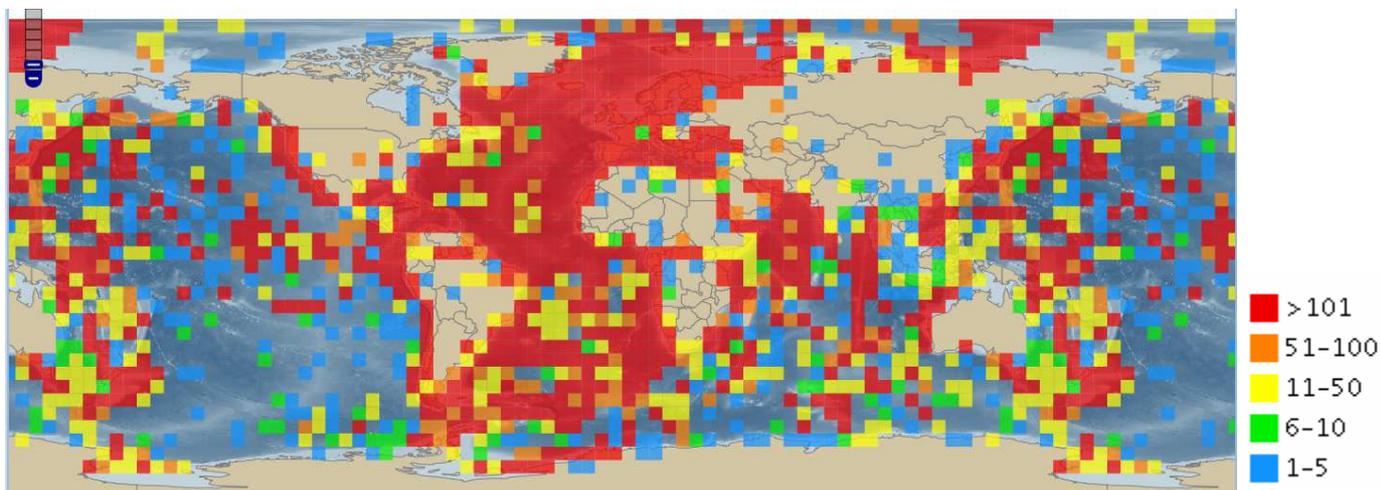


Figure 5. EurOBIS data coverage (years between 1748 and 2009) – extract from OBIS.

### 5.3 ICES

ICES manages marine environmental data covering NE Atlantic, Baltic Sea, Greenland Sea and Norwegian Sea and spanning the years 1877-2012. Those data are organized around specific thematic data portals: oceanographic, contaminants, biological effects and biological community, fish trawl survey, fish predation and historical plankton data.

This paper analyses ICES data coverage separately from the other marine environmental information systems ones. ICES is however expected to be connected to SeaDataNet in a near future and would provide environmental context data to Micro B3 through the SeaDataNet services. ICES biogeographic data are and will be made available through EurOBIS.

The following Figures 6, 7, 8, 9 are the station maps of data in the ICES oceanographic databases (<http://ocean.ices.dk/data/maps/maps.htm>):

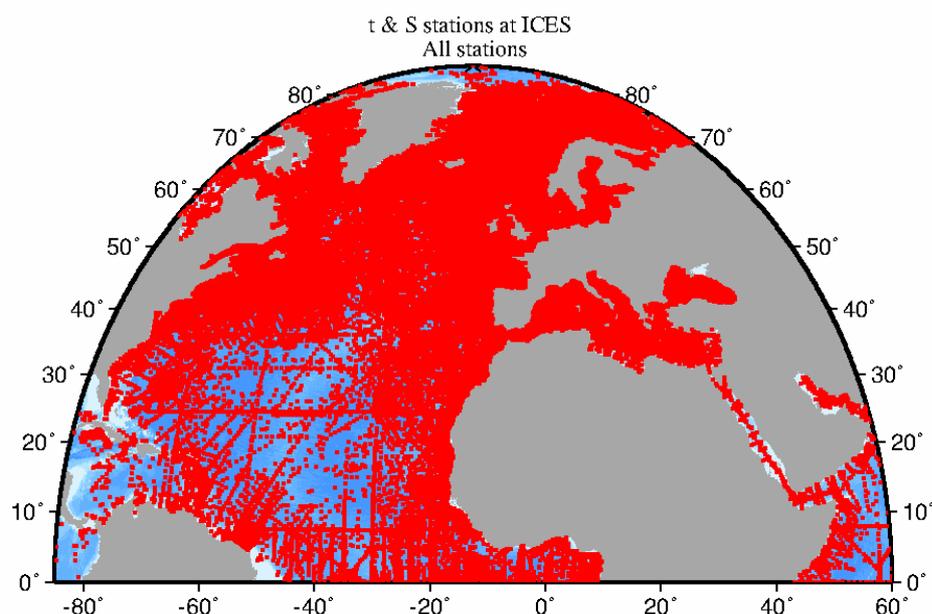


Figure 6. T & S stations at ICES, All stations ([http://ocean.ices.dk/data/maps/All\\_ts.png](http://ocean.ices.dk/data/maps/All_ts.png)).

Figure 7. Nutrient stations at ICES, All stations  
([http://ocean.ices.dk/data/maps/All\\_Nut.png](http://ocean.ices.dk/data/maps/All_Nut.png)).

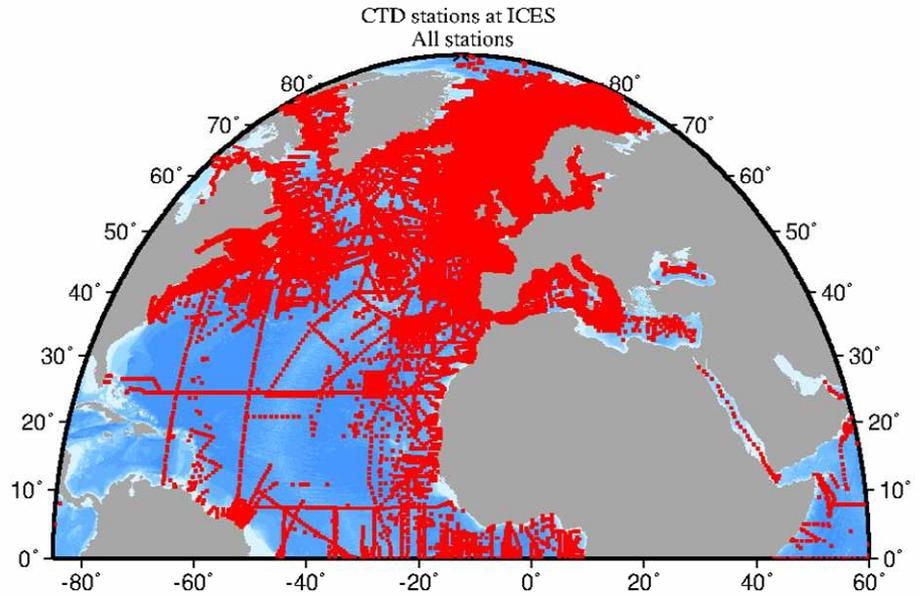


Figure 8. CTD stations at ICES, All stations  
([http://ocean.ices.dk/data/maps/All\\_CTD.png](http://ocean.ices.dk/data/maps/All_CTD.png)).

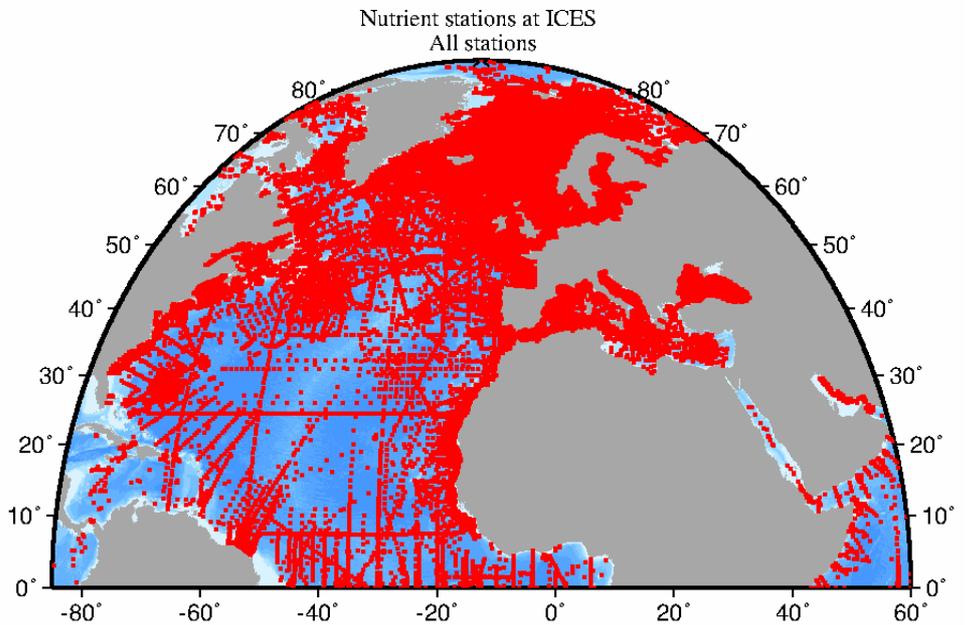
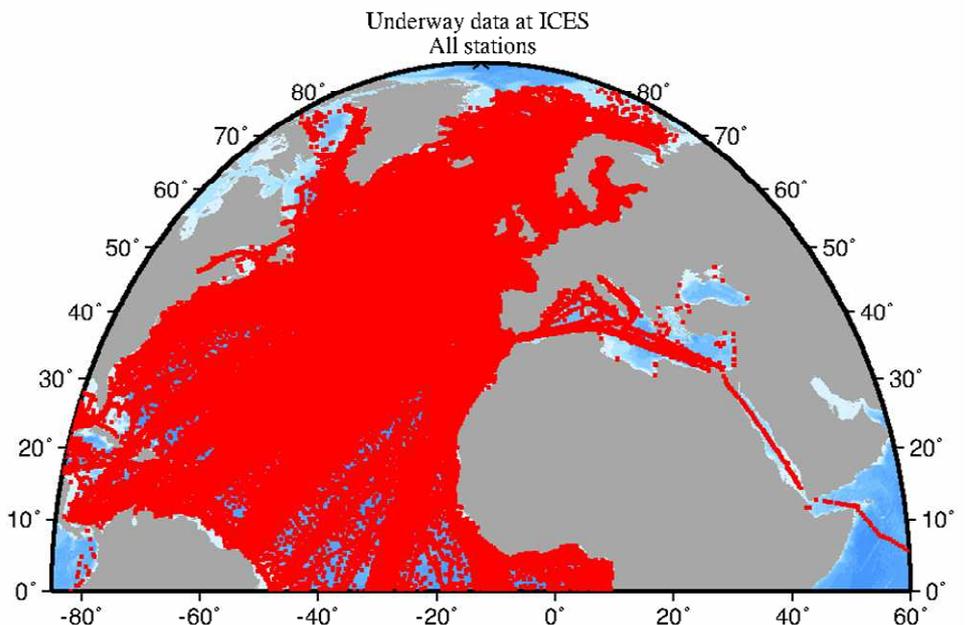


Figure 9. Underway data at ICES, All stations  
([http://ocean.ices.dk/data/maps/All\\_UW.png](http://ocean.ices.dk/data/maps/All_UW.png)).



## 5.4 PANGAEA

PANGAEA manages and can provide to the Micro B3 project environmental and biological data, including an extensive range of parameters describing the life history and vital rates of marine plankton (viruses, bacteria, autotrophic and heterotrophic protists, crustaceans and jellyfish) and microbenthos from contemporary to paleobiogeographic records.

Furthermore PANGAEA will be the receptacle for the Tara Oceans metadata and data and will allow to make them available to the Micro B3 data management system.

As it is the case for ICES, it is planned that PANGAEA will be connected soon to the SeaDataNet infrastructure for giving overview and access to its data sets via the SeaDataNet CDI Data Discovery and Access service. Furthermore, PANGAEA biogeographic data are and will be made available through EurOBIS.

## 6.0 Selection of pilot geographical and temporal sites/areas

### Pilot „temporal“ locations

Based on the partners list of „long-term“ stations the following pilot sites are proposed:

Name	Description	Latitude	Longitude
L4	English Channel (UK)	50.25	-4.22
Helgoland	North Sea (Germany)	54.19	7.90
Naples	Naples (Italy)	40.81	14.25
Heraklion-M3A	Cretan Sea (Greece)	35.78	24.92
Blanes	Spain	41.67	2.80
Matis	Faxafloi (Iceland)	64.08	-22.13

For these sites the environmental data coverage is guaranteed by SeaDataNet and ICES, as well as by EurOBIS for species data.

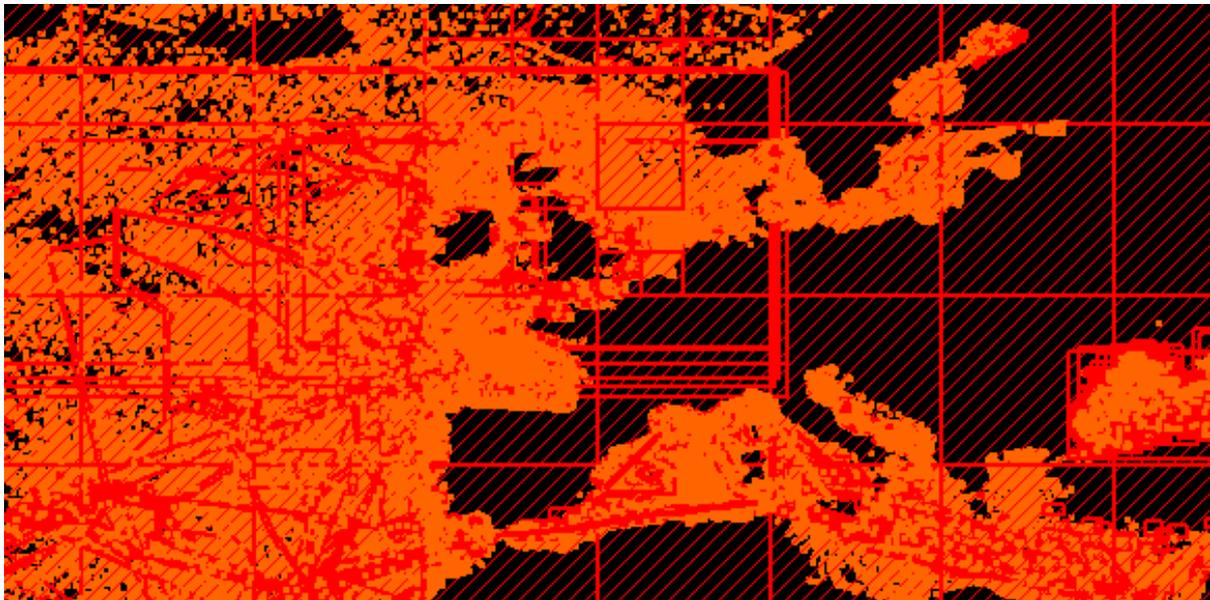


Figure 10. SeaDataNet data coverage (points, tracks and areas) for the timescale 1980-2012.

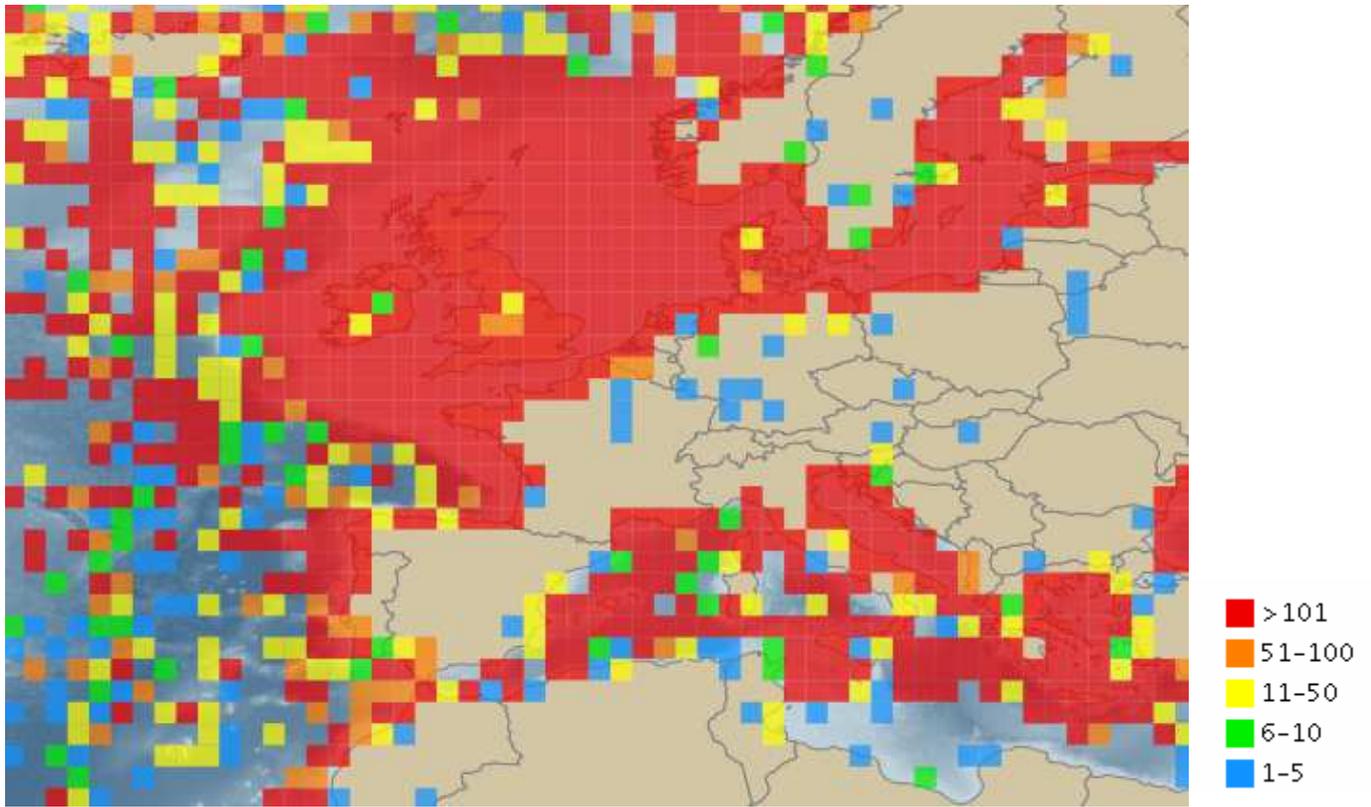


Figure 11. EurOBIS data coverage (years between 1748 and 2009) – extract from OBIS.

### Pilot „geographical“ locations

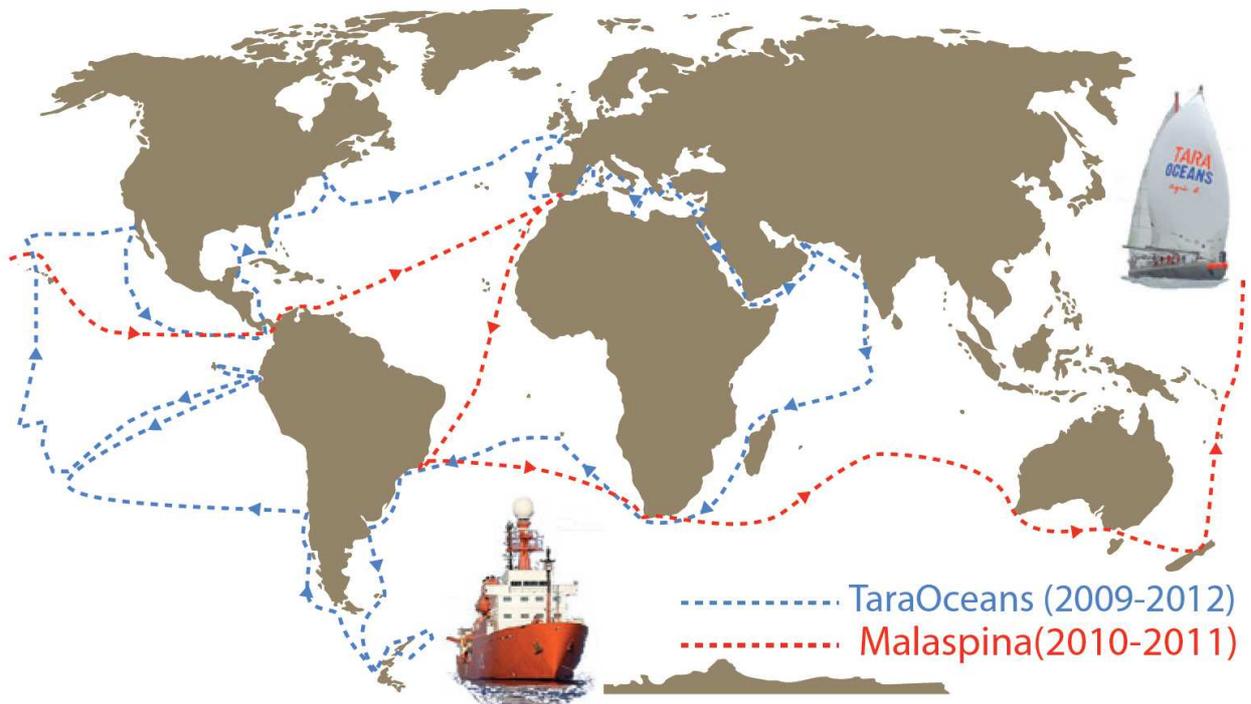


Figure 12. Overview of the Tara Oceans and Malaspina circumnavigation cruises.

When crossing data from the environmental data providers with Tara Oceans and Malaspina expeditions data (see Figures 4-9 and 12), four data „hot spots“ were revealed:

<b>Name</b>	<b>Latitude N</b>	<b>Latitude S</b>	<b>Longitude W</b>	<b>Longitude E</b>
Gibraltar	38	33	-13	-2
South Africa	-30	-40	10	30
Hawai	30	20	-170	-150
Atlantic NE	50	25	-40	-10

Those locations were confirmed by WP2 and WP6 to be of relevance for the MicroB3 Use Cases in light of the samples and variables analyzed through their participants work.



## Reference list

Karsenti E, Acinas SG, Bork P, Bowler C, De Vargas C, et al. (2011) A Holistic Approach to Marine Ecosystems Biology. PLoS Biol 9(10): e1001177. doi:10.1371/journal.pbio.1001177

Not F, Le Bescot N, Pesant S, Kandels-Lewis S, Picheral M, et al. (in prep) Tara Oceans expedition: Plankton Sampling Strategy & Methods. PLoS Biol or Nature Methods