Marine Microbial Biodiversity, Bioinformatics & Biotechnology

Grant agreement n°287589

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

Deliverable 4.2

Best practices workshop and e-conference report

Version: 2
Circulated to: Name (Date)
Approved by: Name (Date)

Expected Submission Date: 30.09.2012
Actual submission Date: 12.10.2012

Lead Party for Deliverable: Guy Cochrane, EMBL-EBI, Hinxton, UK

Mail: cochrane@ebi.ac.uk Tel.: +44-1223-492564

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)	

The Micro B3 project is funded from the European Union's Seventh Framework Programme (Joint Call OCEAN.2011-2: Marine microbial diversity – new insights into marine ecosystems functioning and its biotechnological potential) under the grant agreement no 287589. The Micro B3 project is solely responsible for this publication. It does not represent the opinion of the EU. The EU is not responsible for any use that might be made of data appearing herein.

Summary

The EMBL-EBI, Hinxton, UK, organized on July 5th and 6th 2012 the MicroB3-WP4 Sampling Groups Workshop. Participants from thirteen Institutes (BAS, CNRS, CIESM, CSIC, EMBL-EBI, EMPA, HCMR, JacobsUni, MARIS, MARUM, MBA, UOXF, VLIZ) discussed current and best practices in marine sampling. Variation vs. consistency of information components captured at 8 sampling Sites (L4 UK, Roscoff France, Helgoland Germany, Naples Italy, Crete Greece, Blanes Spain, Rothera Antarctic Peninsula and VLIZ Belgium) has been analysed leading to discussions on critical, desirable and useful aspects of a sample and data processing.

Submission and archiving of marine data in oceanographic (SeaDataNet, Pangaea) and genomic (ENA) repositories have been reviewed. Legal aspects of the MicroB3 project objectives were outlined and a draft of the Mediterranean Code of Conduct presented to the representatives of the sampling groups at the meeting.

The Sampling Groups Workshop provided an insight into current sampling practices in place and refined a design of the Sampling Groups Survey aiming to review information components currently being captured at the MicroB3- and external sampling Sites.

Outcomes of both the Workshop and the Survey reported here will be helpful for further development of the MicroB3 Standards and Interoperability structures.

Table of Content

Objectives of the Sampling Groups Workshop and the Sampling Groups Survey

This document reports on the Sampling Groups Workshop, organised by the EMBL-EBI on 5th and 6th of July 2012, and on the Sampling Groups Survey designed following up the Workshop.

The main task of the Sampling Groups Workshop was to develop better understanding of current and best practices in marine sampling.

The aim of the Sampling Groups Survey was to review information components currently being captured at the MicroB3- and external Sampling Sites.

Information gathered from both the Workshop and the Survey will facilitate assessment of best sampling practices and building of consensus between Sampling Groups needed to evaluate parameters describing marine microbial samples as critical (i.e. mandatory), desirable (i.e. recommended) and useful (i.e. optional).
It will also help understanding and harmonization of oceanographic and genomic data flow from sampling events to the repositories and ultimately to the marine science community.

Sampling Groups Workshop Participants

In total, twenty two Institutes were contacted and invited for the Sampling Groups (SG) Workshop. All invited representatives are listed below with their contact details, their Institute affiliation and, where appropriate, grouped according their affiliation to a sampling Site. Participants of the SG Workshop are highlighted in bold.

Blanes Bay, Spain
Carlos Pedrós-Alió - cpedros@icm.csic.es - CSIC
Josep M. Gasol - pepgasol@icm.csic.es - CSIC
Bibiana G. Crespo - bibiana@icm.csic.es - CSIC

Crete, Greece
George Kotoulas - kotoulas@her.hcmr.gr - HCMR

Hegoland, Germany
Antje Boetius - aboetius@mpi-bremen.de -AWI
Julia Schnetzer - jschnetz@mpi-bremen.de - JacobsUni
Alban Ramette - aramette@mpi-bremen.de - AWI

Island
Viggó Thór Marteinsson - viggo.th.marteinsson@matis.is - MATIS

Naples, Italy

Adriana Zingone - zingone@szn.it - SZN

Roscoff, France
Chris Bowler - cbowler@biologie.ens.fr - CNRS
Daniel Vaulot - vaulot@sb-roscoff.fr
Fabrice Not - not@sb-roscoff.fr

Rothera, Antarctic Peninsula
Melody Clark - mscl@bas.ac.uk - BAS

VLIZ, Belgium
Simon Claus - simon.claus@vliz.be - VLIZ
Klaas Deneudt - klaas.deneudt@vliz.be - VLIZ

Western English Channel (WCO), United Kingdom
Ian Joint - ian.joint@gmail.com - MBA
Declan Shroeder - dsch@mba.ac.uk - MBA

Michele Barbier - mbarbier@ciesm.org CIESM

Dick Schaap - dick@maris.nl MARIS

Stephane Pesant - spesant@marum.de - MARUM, UniHB, Pangaea

Frank Oliver Gloeckner- fog@mpi-bremen.de - JacobsUni

Johanna Wesnigk - j.wesnigk@empa-bremen.de EMPA

Dawn Field - dfield@ceh.ac.uk - UOXF

Peter Sterk - sterk@ebi.ac.uk - UOXF

Guy Cochrane - cochrane@ebi.ac.uk - EBI

Petra ten Hoopen - petra@ebi.ac.uk - EBI

Stephane Riviere - sriviere@ebi.ac.uk - EBI

Charles Cook - ccook@ebi.ac.uk - EBI

Sinan Husrevoglu - sinan.husrevoglu@mam.gov.tr -TUBITAK

Patric Wincker - pwincker@genoscope.cns.fr - Genoscope

Linda Amaral Zettler - amaral@mbl.edu - MBL Woods Hole, USA

Gilbert Maudire - gilbert.maudire@ifremer.fr - IFREMER

Fergal O'Gara - f.ogara@ucc.ie - Biomerit

Michail Yakimov - michail.yakimov@iamc.cnr.it IAMS

Neil Holdsworth- NeilH@ices.dk - ICES

Sampling Groups Workshop Agenda

5th July afternoon (EBI - Courtyard Room)

12.00 – 1.00 – lunch
1.00 – 1.15 – introduction to the SG Workshop
1.15 – 1.30 – WP4: strategy and the story so far
1.30 – 3.00 – OSD pilot review (Dawn Field)
3.00 – 3.30 – sampling practices consensus and differences (Petra ten Hoopen, Stephane Riviere)
3.30 – 4.00 – coffee break
4.00 - 4.20 – legal aspects (Michele Barbier)
4.20 – 6.30 – sampling practices and methodology discussion (Petra ten Hoopen, Stephane Riviere and others)
7.30 pm – dinner at Hinxton Red Lion

6th July morning (EBI - C209 Room)

9.00 – 10.00 – review of repositories for oceanographic and genomic data
9.00 – 9.20 – SeaDataNet (Dick Schaap)
9.20 – 9.40 – Pangaea (Stéphane Pesant)
9.40 – 10.00 – ENA (Guy Cochrane + Rajesh Radhakrishnan)
10.00 – 10.30 – sampling practices and methodology discussion (Petra ten Hoopen, Stephane Riviere and others)
10.30 – 11.00 – coffee break
11.00 – 11.45 – planning and next steps
11.45 – 12.00 – summary and wrap-up
12.00 – 13.00 – working lunch

First day

Guy Cochrane, leader of the MicroB3 – WP4, welcomed all to the Sampling Groups Workshop and invited participants of the meeting to the "tour de partners", where everybody shortly introduced himself.

WP4: strategy and the story so far

Petra ten Hoopen (EMBL-EBI) briefly outlined the overall objectives and strategy of the MicroB3 Work Package 4 – Standards and Interoperability (Figure 1). The talk also summarized prototype Use Cases, identified with help of the MicroB3 Consortium partners, and explained transformation of the Use Cases into the scientific and legal component of the MicroB3 Candidate Checklist, which was also presented to the participants of the meeting.
Detailed description of the MicroB3 prototype Use Cases and the Candidate Checklist can be found in the *Use Case Document* and in the MicroB3 *Deliverable D4.*

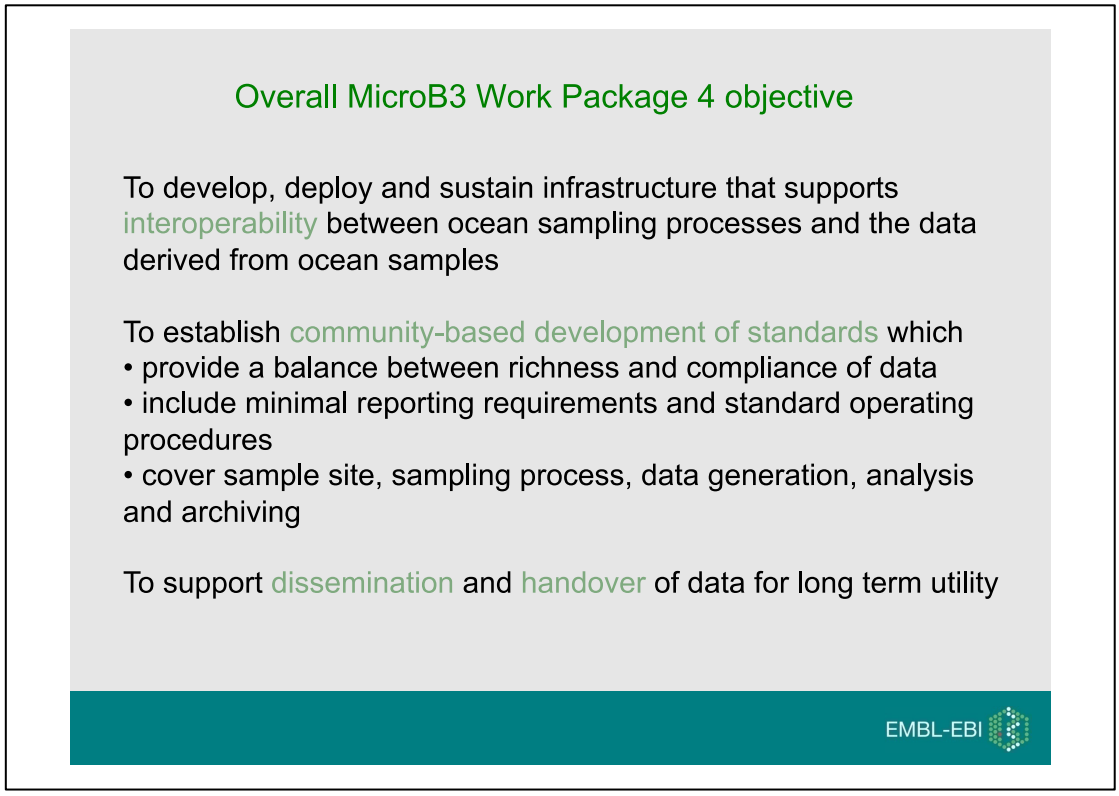

Figure 1: Standards and Interoperability Work Package objectives.

OSD pilot review

In the following session Dawn Field (UOXF) reviewed the Ocean Sampling Day concept and progress, and summarized results of the pilot OSD taking place on 20th June 2012.

The Ocean Sampling Day (OSD) is a simultaneous sampling campaign of the world's oceans aiming to reveal marine microbial diversity and identify novel ocean-derived biotechnologies.
The OSD provides a unique opportunity to study spatial distribution of marine microbial diversity in fixed time; unlike LTER monitoring sites or sampling cruises that allow insight into a temporal biodiversity distribution in a fixed or variable geographic coordinates, respectively.

The OSD 2012 pilot study has been put together in collaboration with the Genomic Observatories Network (GOs Network, http://www.genomicobservatories.org), the Earth Microbiome Project (EMP, http://earthmicrobiome.org) and the Global Genome Initiative (GGI, http://www.mnh.si.edu/ggi).

Twenty sampling Sites carrying out sustained research contributed to the OSD 2012 pilot, which focused on bacterial diversity of the water column. Ocean samples with metadata compliant with the GSC's MixS standards will be sequenced by the EMP according to EMP protocols and bio-archived at the GGI. EMP offered to sequence 10.000 samples at no costs and there is currently no European alternative to that. In the future, sequencing of the OSD samples should be a collaborative initiative of sequencing centres.

Discussions during this session addressed a number of issues

- Tentative and final destination of sample metadata and sequenced data

  All data from the OSD 2012 pilot study will be returned to the submitting institutions but will also remain at the EMP.
  All OSD sequence data should ultimately be archived at the INSDC via the metagenomics/metatranscriptomics portal being developed at the EMBL-EBI, Hinxton, UK.
  All OSD oceanographic data should ultimately be archived at the SeaDataNet via the submission portal of the Pangaea database.

- Availability of the OSD sample data and sequence data

  It has been agreed that the OSD sample metadata and sequence data should be made public as soon as necessary validation steps have been completed and there should be no moratorium on the data.

- Growing interest in the OSD

  Since there is a growing number of sampling Sites interested in participating in the OSD it has been suggested that there might be needed regional coordinators for the OSD, for instance, a coordinator for the Mediterranean Sea or for the Pacific Ocean.

A blog has been launched regarding the OSD, available at http://oceansamplingday.blogspot.co.uk/2012/06/ocean-sampling-day-blog-launched.html. MicroB3 consortium partners and the OSD participating sampling groups who have been granted an access can obtain more information regarding the OSD at the *OSD Google Docs* that provide details on the OSD sampling Sites, network of people, data policy and OSD 2012 pilot samples and metadata.

- Next OSD pilots prior to the main OSD in June 2014

  The following OSD pilot studies will focus on comparing biodiversity indexes between 2 datasets produced in summer and winter solstices with the assumption that lower latitudes will show higher biodiversity.

Legal aspects of the OSD

In this session Michele Barbier (CIESM) briefly mentioned the Nagoya Protocol – Convention on Biological Diversity. Reasons for a need of this Access and Benefits Sharing (ABS) protocol being

- ensuring of fair share of benefits
- promoting transfer of technologies
- enhancing training, research and development
- conservation and sustainable use of genetic resources

Michele also reminded to the participants of the meeting minimal requirements for ABS

- providing provenance of the Marine Genetic Resource (MGR)
- identifying the MGR
- traceability of the MGR, the provider, third parties and recipients
- defining PIC (Prior Informed Consent), MAT (Mutually Agreed Terms)
- demonstrating compliance with domestic ABS

In the second half of the talk Michele introduced a CIESM proposal of the Mediterranean Code of Conduct, which is a morally binding and ABS supporting agreement to be signed by all OSD participants.

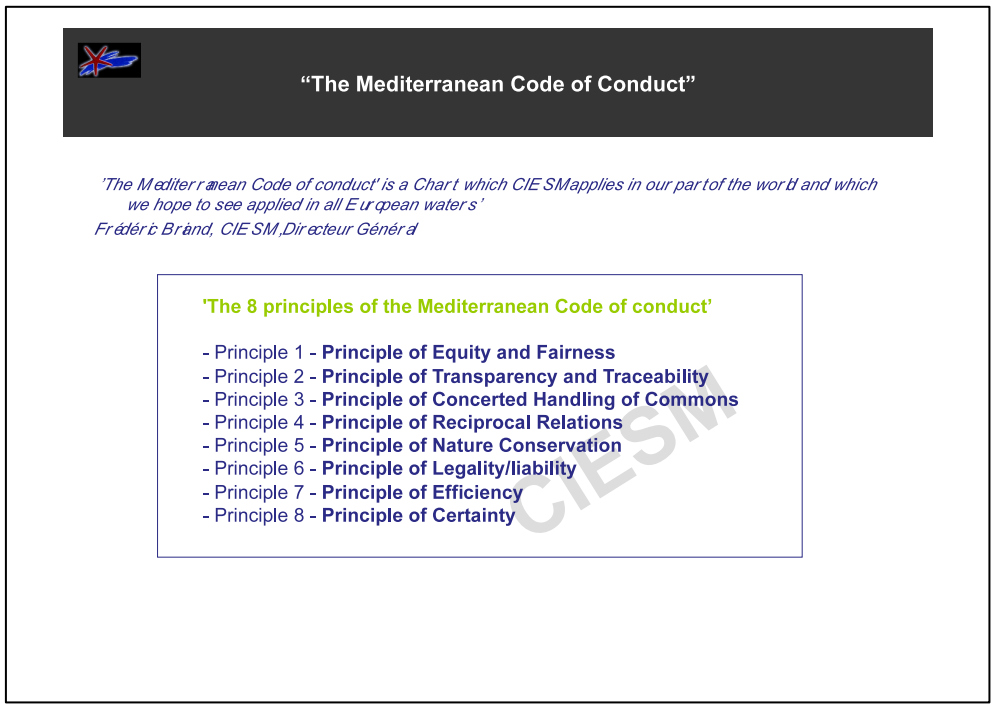

Figure2: A draft of eight Principles of the Mediterranean Code of Conduct.

The Mediterranean Code of Conduct is a set of Principles (Figure 2) that gathers and proposes behavioural guidelines to improve research relationships and/or trade for the use

of MGRs, from transfer and analysis of the MGRs to data management, publications, intellectual property rights and patents. There are no legal obligations to enforce the Code of Conduct and no penalty for it's violation but it is morally binding transient instrument that engages collaborators to meet the Principles.

Sampling practices and methodology discussion

In this section of the SG Workshop we wanted to develop better understanding of working practices in sampling at different ocean sampling Sites, find consensus and discuss differences between them. Furthermore, this session provided opportunity for sampling Sites to highlight specific aspects of their sampling and express opinion on standardization of sampling methods.

Sampling practices and methodology discussions have been structured into several topics

- sample capture
- sample processing and archiving
- data generation and analysis
- specific aspects of sampling Sites

Sampling Sites invited to the SG Workshop have been asked prior to the SG Workshop to provide a spreadsheet with parameters routinely recorded by their sampling group.
Routinely captured parameters from six sampling Sites (Blanes Bay, Hegoland, L4, Naples, Roscoff and Rothera) have been compared and mapped to mandatory GSC MIMARKS water sample fields and to information components requested to be recorded by the sampling Sites during the OSD 2012 pilot study (Annex 1 – Sampling practices consensus).

Several parameters are routinely captured at each sampling Site, such as salinity or temperature, but there is also a great variability between the metadata sets, caused possibly by the fact that some parameters are obviously routinely captured at each site but were not reported in the spreadsheet, such as latitude and longitude. High variability may also reflect differences in the interpretation of the request, i.e. some sites provided more comprehensive information on all measured parameters while others listed only parameters for one representative sample.
We have therefore decided to address the questions of a sample capture and sample processing in more consistent way and agreed on a need for a Sampling Groups Survey. More details on the questionnaire, its results and analysis are discussed in the last two sections of this report.

Participants of the meeting also discussed how- and to which extend should methodologies be standardised.
Two methodological parameters were agreed to be critical for the minimal reported requirements

- *in situ sample filtering*
- *sample storage conditions*

Providing *a sample replicate number* was generally considered desirable and reporting on other methodologies, such as HPLC, flow cytometry, nutrient analysis or sample contamination issues, were evaluated only as useful.

Further assessment of methodologies during the MicroB3 standards concepts development will be necessary to ensure that methodological information is consistently described, includes sampling in both environments, water and sediment, and covers the whole microbial taxonomic spectrum.

In the next part of this session we disucssed a generation and analysis of the MicroB3 data.

Dawn Field suggested a possible MicroB3 data outcome (Figure 3). *Genomic sequences* represent here sequences of specific loci; particularly 16S rRNA gene, 18S rRNA gene, intergenic spacers (ITS); and metagenomics data. While the prokaryotic diversity research benefits mainly from metagenomics data and OTU data matrix (Figure 4), metatranscriptomes analysis and microscopy imaging of microbial structures are essential for eukaryotic studies, for example for diatom biology.
Oceanographic in situ *data* will be provided by the sampling Sites during the OSD while interpolated environmental measurements (ancillary data, climatology information and predictions) can be obtained from the sampling Sites independently of the OSD.

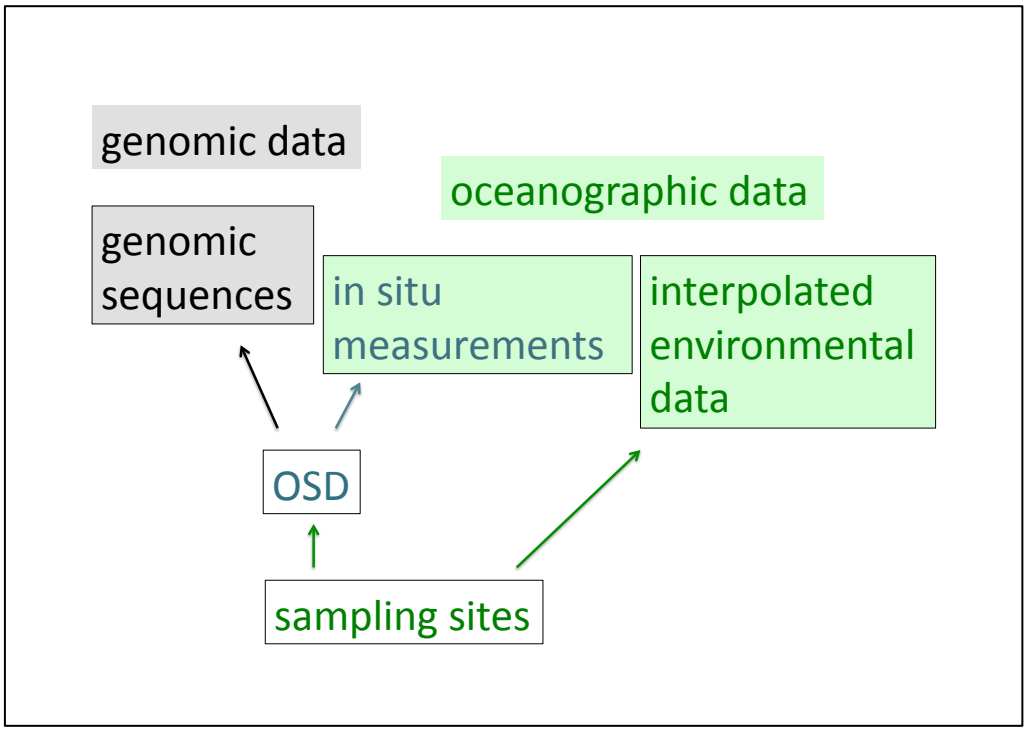

Figure 3: Schematic representation of the suggested MicroB3 data output.

sampling sites

OTU	A	B	C	D
1	30	0	30	0
2	0	0	0	0
3	1	1	0	11
4	0	0	1	1

Figure 4: Example of OTU (operational taxonomic unit) matrix as a representation of prokaryotic biodiversity recorded by individual sampling Sites during the OSD.

In the last part of this session three sampling Sites (Blanes Bay - Spain, VLIZ - Belgium and Western English Channel Observatory United Kingdom) highlighted thier specific sampling practices.

Bibiana G. Crespo introduced Blanes Bay in Spain (Figure 5), a warm, salty and nutrient poor oligotrophic coastal system with episodic intrusions of oceanic waters. Monthly sampling at the Blaney Bay observatory focuses on abundance, activity and diversity of phytoplankton, bacteria, cyanobacteria and viruses.

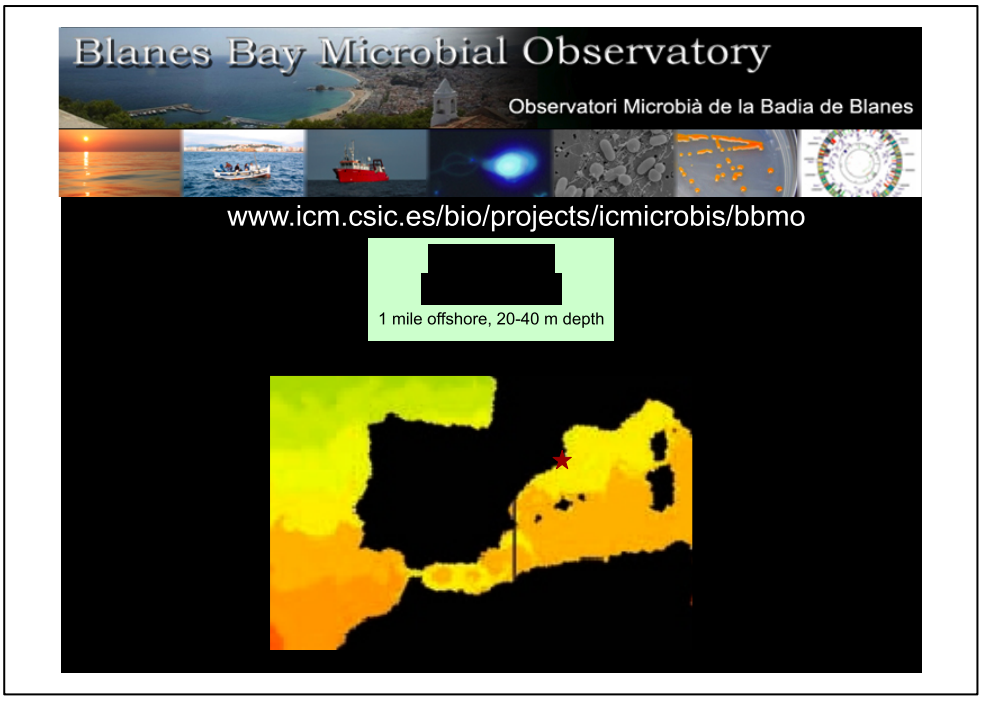

Figure 5: Blanes Bay microbial observatory in Catalonia, Spain.

Simon Claus described instrumental measurements and listed physical samples taken from water and sediment at VLIZ monitoring campaign in Belgium (Figure 6). Simon provided an informative overview of physical, chemical and biological parameters captured at VLIZ, sources and final archiving destinations of the obtained data.

Figure 6: Data captured at VLIZ, Belgium.

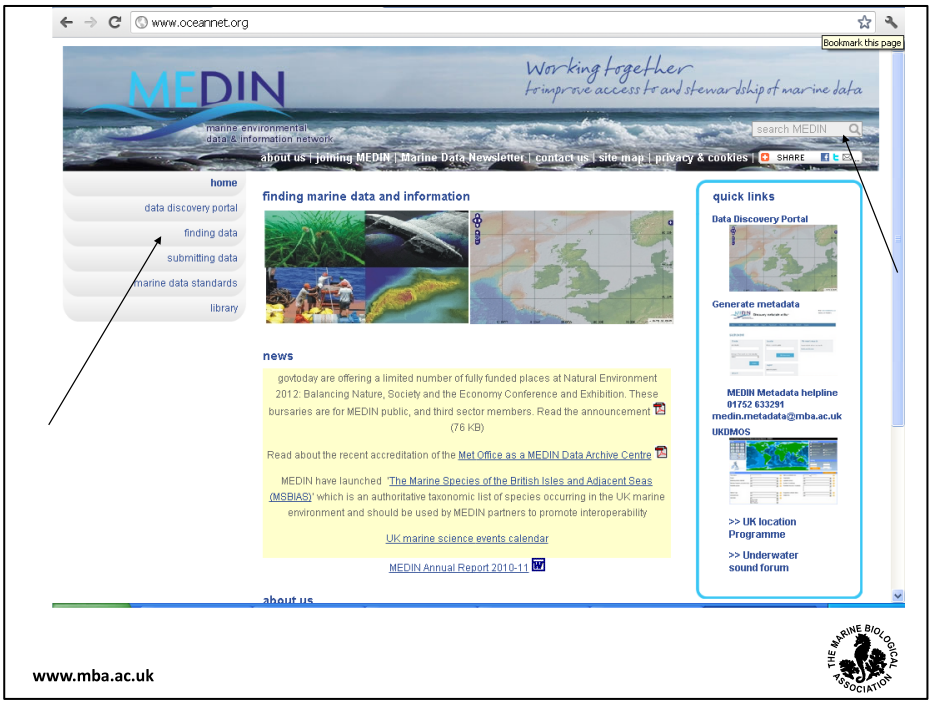

Figure 7: MEDIN is a collaborative effort of public and private sector partners aiming to improve access to and standardization of marine data.

Declan Shroeder introduced the Western English Channel Observatory with it's weekly oceanographic time-series at the coastal L4 sampling station and fortnightly series at the open shelf E1 station, managed by the Plymouth Marine Laboratory and the Marine Biological Association. Declan gave an overview of British oceanographic data archive centres and focused also on efforts of MEDIN (Figure 7) that can be of interest for the MicroB3 standards concepts development.

Second Day

Participants of the SG Workshop reviewed oceanographic and genomic repositories in order to unify terminologies and improve understanding between the genomic and the oceanographic community.

Dick Schaap from the SeaDataNet, http://www.seadatanet.org/, gave in his talk a comprehensive review of the oceanographic data acquisition and outlined a role of the SeaDataNet as a portal with harmonised services, tools, standards and data products (Figure 8). The SeaDataNet is building a European infrastructure for managing marine and ocean data by connecting National Oceanographic Data Centres (NODCs) and oceanographic focal data points from 35 coastal states in Europe. Dick also mentioned the European Marine Observation Data Network (EMODnet), which is an open access network of existing and developing European observation systems covering all European coastal waters, shelf seas and surrounding ocean basins.

Dick reminded that acquiring genomic samples, i.e. ocean samples for genomic research, is only one of many streams of data and samples collected from the oceans and seas.
Dick recommended that genomic samples collection and processing as well as data management are in agreement with protocols and practices agreed by oceanographic data centres that can assure that oceanographic data are stored and documented in a proper way and are accessible via the SeaDataNet infrastructure. Genomic samples should be processed as part of the normal oceanographic data stream but additional layer of genomic data-specific requirements for these samples will assure their compliance with genomic standards and their long-term archiving in repositories for genomic data. Repositories for oceanographic and genomic data will then establish mutual interoperability for serving both communities.

In the next talk Stéphane Pesant introduced the Pangaea, http://www.pangaea.de/ (Figure 9), an information system aimed at archiving, publishing and distributing data associated with pelagic and benthic ecology and paleooceanography. Stéphane reviewed infrastructure of the Pangaea, it's data content, data submission and curation systems, and data dissemination implemented by a public Google-like search of metadata, a data warehouse for extensive downloads and web-services. Editorial system in Pangaea enables to cite each data point with DOI number and allows cross-references with journal articles and other databases.
It has been suggested that there should be one MicroB3-OSD account at the Pangaea and each sample identifier or sampling event identifier issued by the Pangaea should be linked

to accession numbers of the International Nucleotide Sequence Database Collaboration (INSDC).

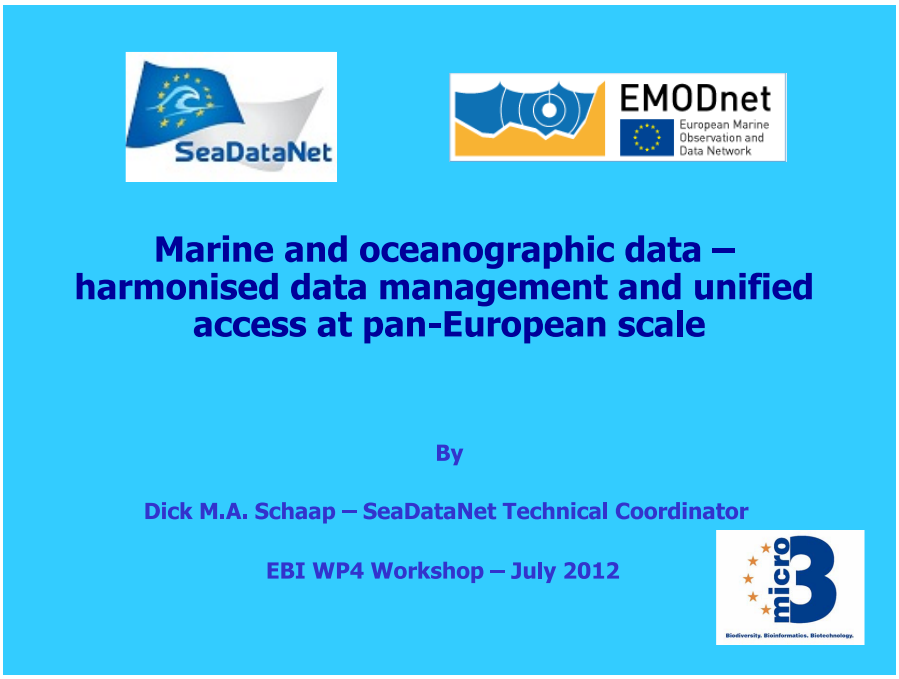

Figure 8: SeaDataNet develops infrastructure and standards for marine data management.

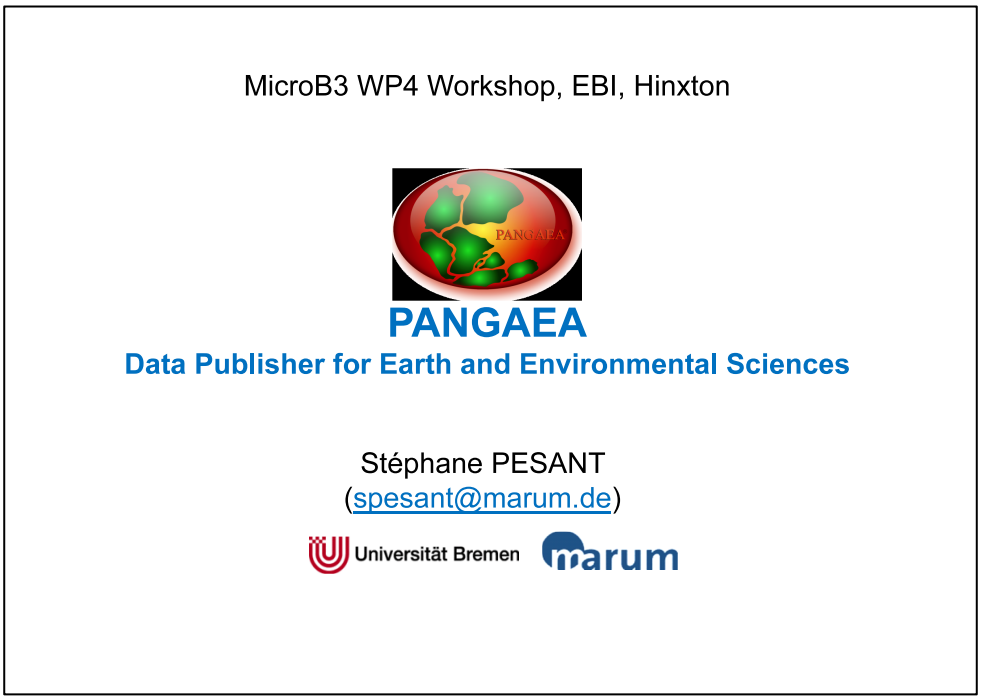

Figure 9: Pangaea is among others the data manager for the TARA Oceans project.

Guy Cochrane reviewed in his talk the European Nucleotide Archive, ENA, http://www.ebi.ac.uk/ena/, a globally comprehensive Archive of nucleotide sequence data and the INSDC partner of the GenBank and the DDBJ. The ENA hosts and provides access to raw reads, assemblies and sequence annotation for individual users of nucleotide data and for downstream databases. Nucleotide sequences are assigned permanent identifiers of the sequences – the INSDC accession numbers. Due to ever growing content of the Archive, including marine data (Figure 10), the ENA is considering a CRAM compression of the next generation sequences and more extensive involvement of domain-specific submission brokers in order to sustain the broad spectrum of services that the ENA offers.

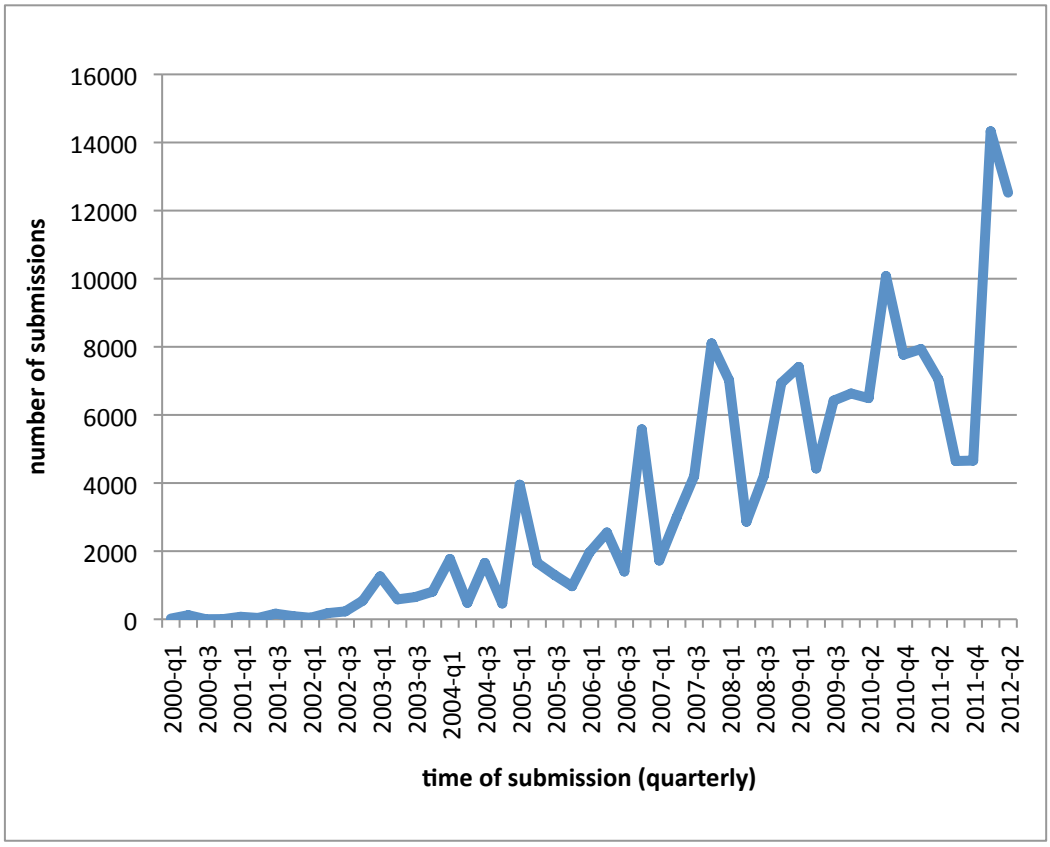

Figure 10: An increase in the number of submissions of marine 16S rRNA gene locus data in a period of the last 12 years.

Guy also drafted in his talk a potential flow of the MicroB3 genomic and oceanographic data (Figure 11a). Sampling groups will submit marine/oceanographic data from their samples to the Pangaea that will pass on a selected subset of key environmental parameters to the ENA and will act as a broker for the SeaDataNet. Sequences of the samples will be submitted to the ENA that will archive and display the sequences with the selected subset of environmental parameters provided by the Pangaea. The ENA and the Pangaea will establish mutual links between the sequence data and marine/oceanographic data. Sequence data from the ENA will flow, via the metagenomic portal, together with the comprehensive marine/oceanographic data from the SeaDataNet to Bremen (MEGX/MEGDB). The MicroB3 infrastructure will provide the genomic data in their environmental context to the broader marine community.

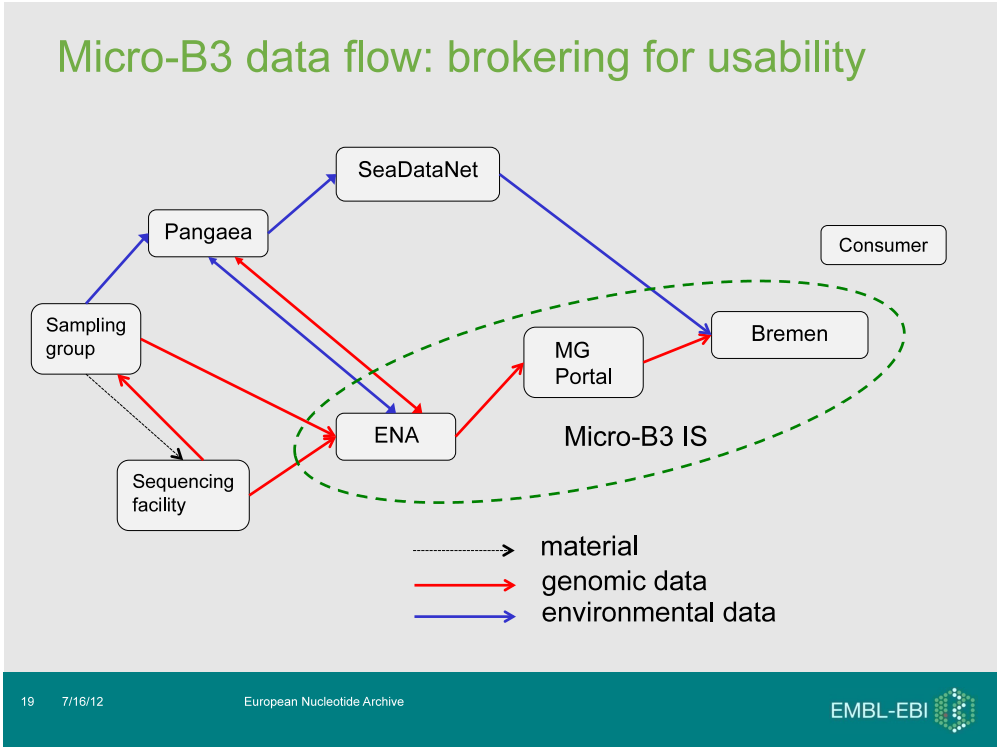

Figure 11a: Illustration of data exchange among oceanographic and genomic repositories suggested at the Sampling Groups Workshop.

In discussions with the MicroB3 partners following up the meeting we have further defined the MicroB3 data flow from the perspective of the user, i.e. the perspective of the sampling groups and the consumer of all data available via the MicroB3 infrastructure. The current proposal of the data flow is represented in the Figure 11b overleaf.

At the end of the meeting Petra ten Hoopen (EMBL-EBI) summarized outcomes from the SG Workshop and next steps of the MicroB3 - WP4 relevant to the sampling groups.

The SG Workshop

- contributed to better understanding between communities, i.e. biologists building genomic collections vs. oceanographic scientists
- revealed consistent and variable aspects of sampling among groups that are further addressed in the Sampling Groups Survey
- reviewed oceanographic and genomic repositories contributing to the MicroB3 infrastructure and their data submission systems that will be used by the OSD participants

Next steps following up the SG Workshop

- the WP4 team designs a Survey of current sampling practices at marine microbial sampling Sites
- information components identified in the Survey will be mapped against the MIxS genomic standards and against existing oceanographic standards
- MicroB3 standards concepts will be drafted and tested against real data sets using identified prototype Use Cases
- the Ocean Sampling Handbook will be assembled with help of sampling groups and other stakeholders

All presentations from the SG Workshop can be found at the MicroB3 wiki page (https://colab.mpi-bremen.de/micro-b3/trac/wiki/WorkPackages/Wp4).

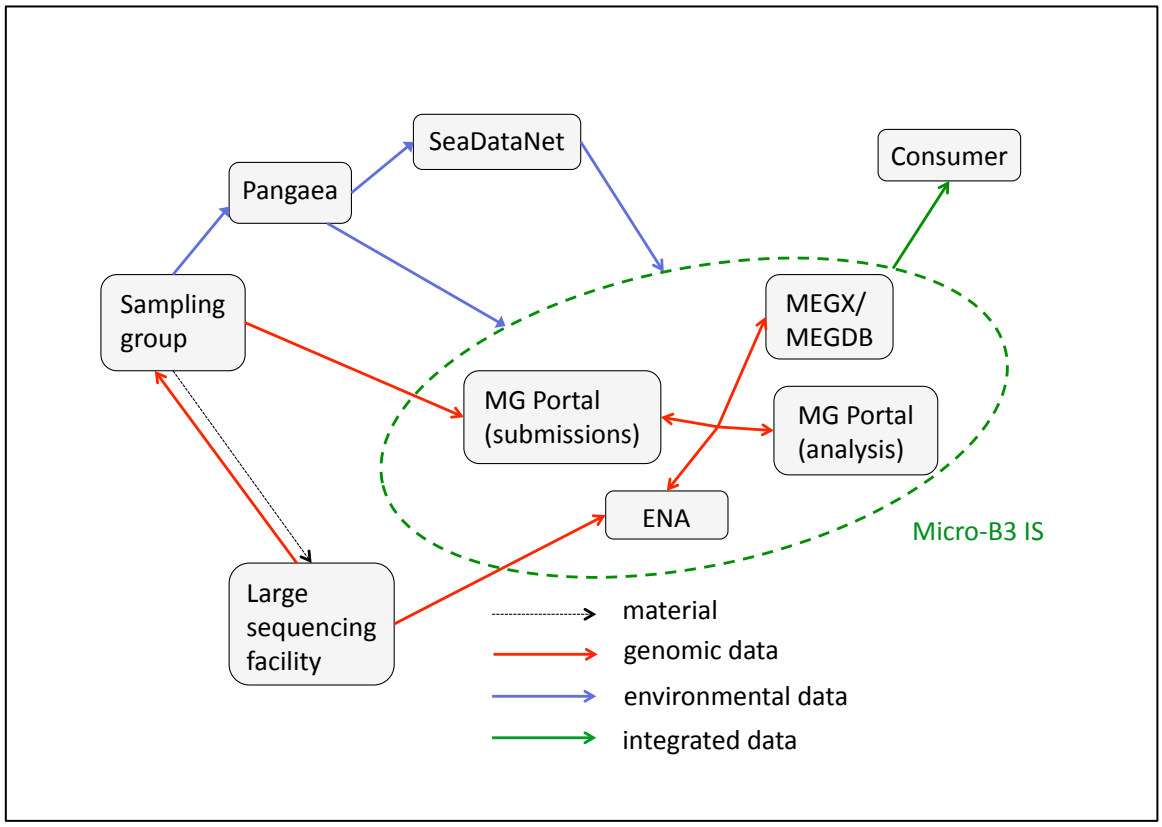

Figure 11b: The currently proposed scheme of the MicroB3 data flow from the perspective of a user.

Sampling Groups Survey

The Sampling Groups Survey has been designed as an e-communication following up the SG Workshop in order to review in a consistent way current sampling practices at sampling Sites either involved in the MicroB3 project or participating in the OSD 2012 pilot study. The Survey focuses on scientific aspects of sampling. Administrative and legally-relevant parameters, such as a project description, a cruise name or the leading scientist contact details, are not addressed in this Survey and will be considered in the collaboration with the MicroB3 Work Package 8.

Information elements for the Survey have been gathered using the NERC vocabulary server (http://www.bodc.ac.uk/products/web_services/vocab/), the SeaDataNet links to metadata services and the oceanographic peer-reviewed literature. A draft of the Survey has been consulted with oceanographic experts (Declan Shroeder – WCO, UK and Bibiana G. Crespo – Blanes Bay, Spain) and the questionnaire amended based on their feedback.
Subsequently, the Survey was designed using the SurveyMonkey software (http://www.surveymonkey.com/), which allows an extensive analysis of responses.
The Survey consists of 18 sections covering broad spectrum of sampling aspects. Each section contains up to 3 questions. Majority of the questions are in the format of a rating scale. Each parameter should be evaluated as *critical*, *desirable*, *useful* or *not applicable* according to the relevance to sampling practices at each sampling Site. Several questions are in the format of a multiple choice where a multiple answer is possible. Each section also gives sampling groups the opportunity to specify additional parameter, which are not included in the Survey and they consider important for their sampling practices.

The Survey has the following sections

- sampling Site name
- descriptive parameters
- sample-related parameters
- meteorology
- sea state, currents and fluxes
- optical parameters
- physical, geophysical and sediment parameters
- nutrients and dissolved gasses in the water column
- carbon organic and inorganic
- nitrogen and phosphorus
- other chemical parameters
- biochemical parameters
- biological parameters
- pigments
- rate and damage measurements
- imaging
- methodology
- data archiving
- comments

The Sampling Groups Survey, available at http://www.surveymonkey.com/s/D58SNP8 , has been sent to the Sampling Groups listed below (Table 1). Sampling groups that completed the questionnaire are highlighted in bold green.

sampling Site	**contact name**	**contact details**
Blanes Bay, Spain	Bibiana Crespo	bibiana@icm.csic.es
Helgoland, Germany	Julia Schnetzer	jschnetz@mpi-bremen.de
Rothera, Antarctica	Melody Clark	mscl@bas.ac.uk
Naples, Italy	Adriana Zingone	zingone@szn.it
Roscoff, France	Daniel Vaulot	vaulot@gmail.com
WCO – L4, United Kingdom	Declan Shroeder	dsch@mba.ac.uk
Crete, Greece	Georgios Kotoulas	kotoulas@her.hcmr.gr
Iceland	Viggó Þór Marteinsson	viggo@matis.is
VLIZ, Belgium	Simon Claus	simon.claus@vliz.be
Thames, United Kingdom	Dan Read	dasr@ceh.ac.uk
Gullmarsfjord, Sweden	Maria Asplund	maria.asplund@gu.se
Banyuls, France	Ian Salter	ian.salter@obs-banyuls.fr
Villefranche, France	Maria Luiza Pedrotti	pedrotti@obs-vlfr.fr
Churchill, Canada	LeeAnn Fishback	fishback@churchillscience.ca
Moorea, French Polynesia	Neil Davies	neiltahiti@gmail.com
BATS, USA – Bermuda	Stephen Giovannoni	steve.giovannoni@oregonstate.edu
SPOTS, USA – San Pedro	Jed Furhman	fuhrman@usc.edu
HOTS, USA – Hawaii	Ed Delong	delong@mit.edu

Table 1: A list of sampling Sites contacted to complete the Sampling Sites Survey

Results and analysis of the Sampling Groups Survey

Results of the SG Survey will provide a solid base for an assessment of the current sampling practices and will facilitate further development of the MicroB3 standards concepts. An analysis of the survey will be helpful in a decision making on how and to which extend we shall aim to standardize marine sampling methodologies.

From the 20 sampling Sites that we have contacted 9 Sites completed the Survey, i.e. 45%. All of the Sites are sampling stations but 2 also have cruises. 6 Sites measure time series, 3 have experience in profile measurements and 1 in sampling trajectories. 2 sites take samples from the benthic zone.

According to all respondents the Survey covered all aspects relevant to the sampling at their Site.

From all parameters questioned in the Survey we have selected those **sampling parameters valued highly (as critical or desirable for sampling) by at least 40% of respondents**. These parameters are grouped according to the Survey section they appeared in and presented in Figure 12-1 till Figure 12-12.

- **Descriptive parameters**

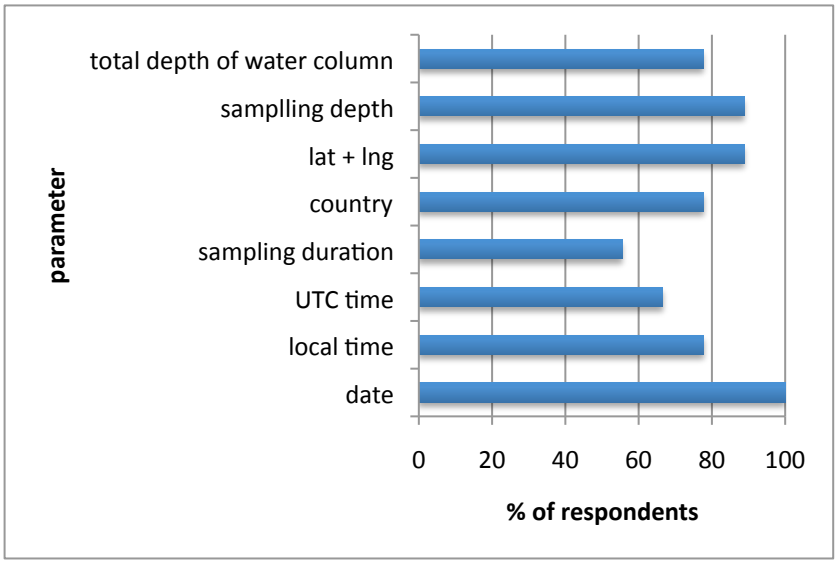

Figure 12-1: Descriptive parameters evaluated by at least 40% of sampling Sites as critical or desirable for their sampling practices.

- **Sample-related parameters**

Sample volume varies in the range 0.1L till 20L, where volume of 1L seems to be a safe minimum for a biological analysis.
Mostly polycarbonate or Sterivex filters are used.
Filter pore size varies in range 0.02 µm till 20 µm, where 3µm is in some cases used for prefiltering and 0.22µm for filtering.

Samples are stored at -20 °C and in case of an RNA analysis, or where possible, at -80 °C.

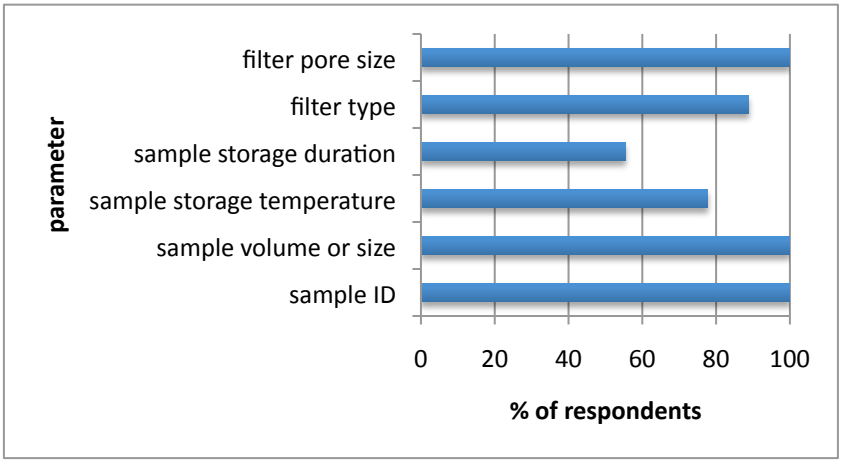

Figure 12-2: Sample-related parameters evaluated by at least 40% of sampling Sites as critical or desirable for their sampling practices.

- **Sea state, currents, fluxes and meteorology**

None of the respondents appreciates recording of currents or fluxes.
The sea state parameter – **tidal stage** – was evaluated as critical by 2 sampling Sites.
Only two meteorological parameters appeared significant, **air temperature** and **wind speed**. Meteorological data are mostly taken from ancillary observations (weather buoys or weather stations).

- **Optical parameters**

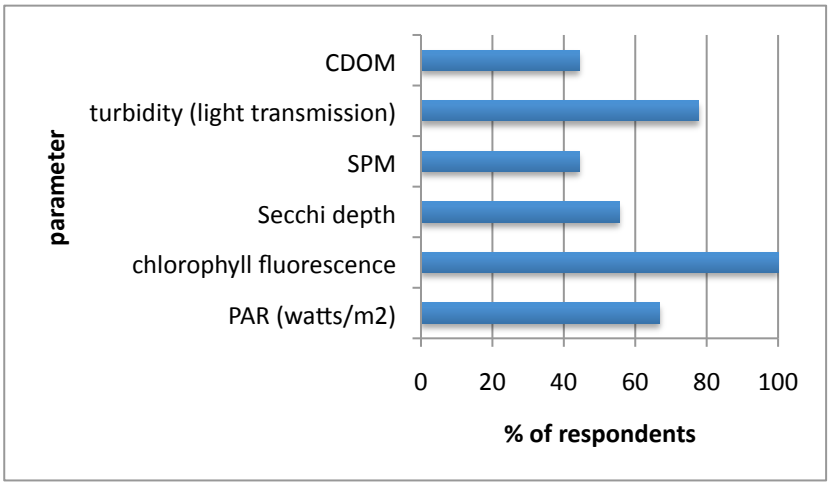

Figure 12-3: Optical parameters evaluated by at least 40% of sampling Sites as critical or desirable for their sampling practices; CDOM – colored dissolved organic matter, SPM – suspended particulate matter.

- **Geophysical and sediment parameters**

One group sampling in sediments found **bathymetry** and sediment parameters listed in the Survey, i.e. **sediment type**, **particle classification**, **porosity**, **pore water content,** critical for their sampling. The second group with experience in sediment sampling evaluated bathymetry as useful and two parameters, sediment type and particle classification, desirable for their sampling.
Further consultations with Sites sampling in sediment will be necessary to establish information elements that can correctly describe sampling in the benthic zone.

- **Physical parameters**

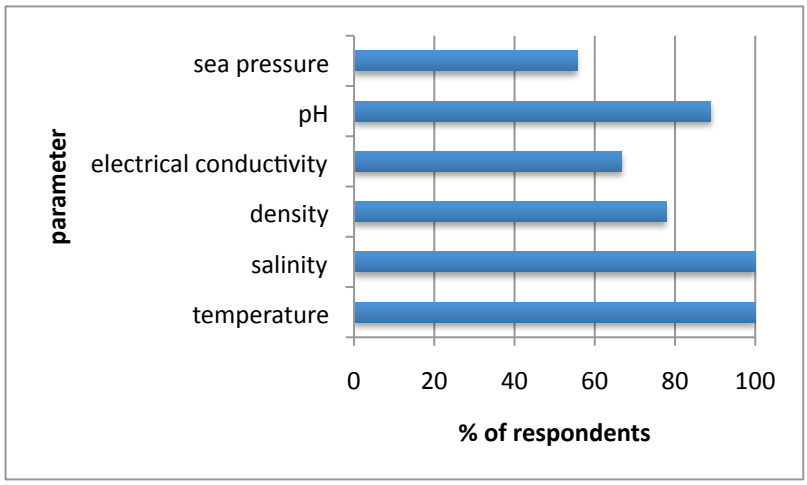

Figure 12-4: Physical parameters valued by at least 40% of sampling Sites as critical or desirable for their sampling practices.

- **Dissolved gasses and nutrients in the water column**

Oxygen and **carbon dioxide** are the only widely measured dissolved gasses.

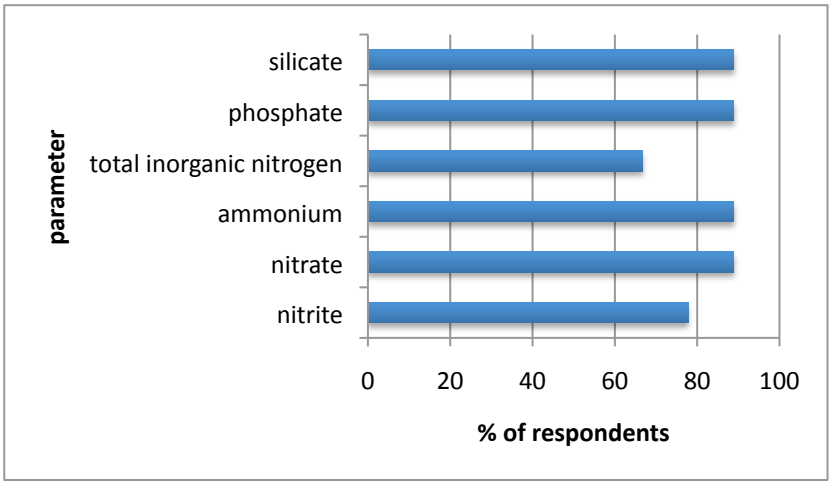

Figure 12-5: Nutrients evaluated by at least 40% of sampling Sites as critical or desirable for their sampling practices.

- **Carbon organic and inorganic, nitrogen and phosphorus**

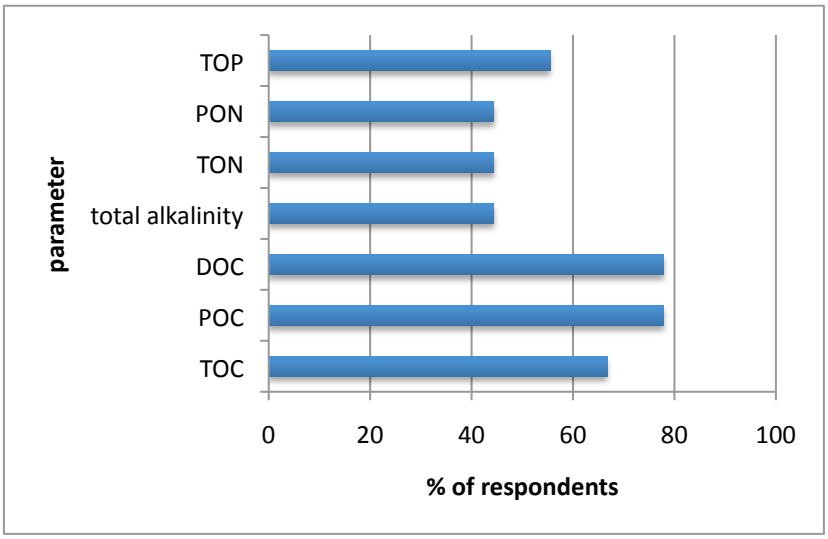

Figure 12-6: Total (T), dissolved (D) and particulate (P) organic carbon (OC), inorganic carbon, nitrogen (N) and phosphorus (P) evaluated by at least 40% of sampling Sites as critical or desirable for their sampling practices.

- **Other chemical elements and ions**

None of the respondents values highly concentration measurements of chemical elements or ions.

- **Biochemical parameters**

8 sampling Sites extract nucleic acid from the residue on the filter, 1 site uses the filtrate and 2 sites take both fractions.

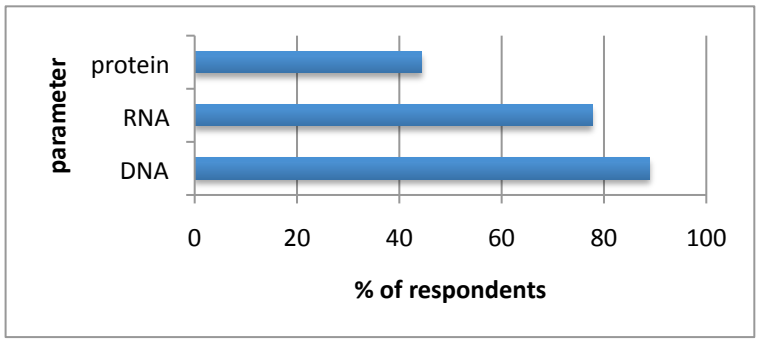

Figure 12-7: Biochemical parameters evaluated by at least 40% of sampling Sites as critical or desirable for their sampling practices.

- **Biological parameters**

Benthic zone is not very well represented here. Only one group, sampling in sediments, analyses all taxonomic groups (viruses, bacteria, phytoplankton, zooplankton) in benthic zone, two other groups focus only on bacteria in this zone.

Pelagic zone has a better representation.

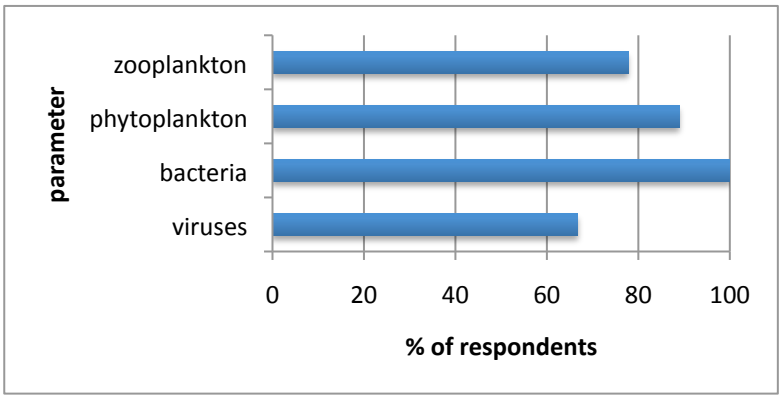

Figure 12-8: Taxonomic groups sampled for in the pelagic zone by at least 40% of respondents.

From all evaluation aspects of the sampled communities of marine microorganisms only 2 Sites found critical the analysis of **the volume in the water column** and 1 group the analysis of **the surface area in the water column**. Recording of **the egg production** is for two groups critical and for one group desirable. Other evaluation aspects of the sampled organisms, which are recorded more frequently, are presented in the Figure 12-9.

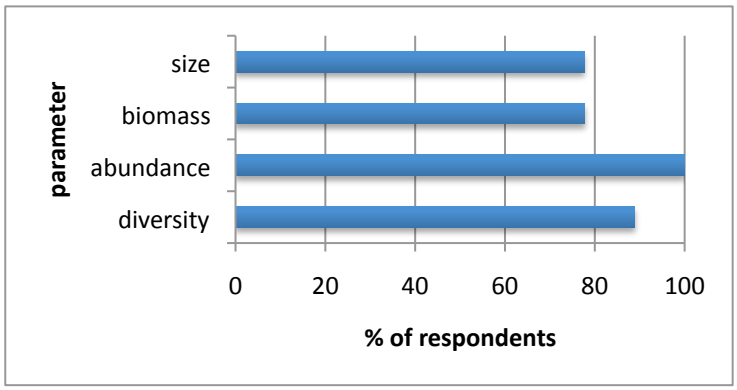

Figure 12-9: Evaluation aspects of the sampled organisms assessed by at least 40% of the Survey respondents as critical or desirable for their sampling practices.

- **Pigments**

All 9 respondents capture pigments during their sampling from the water column, where 2 of them analyse pigments also from the sediments.

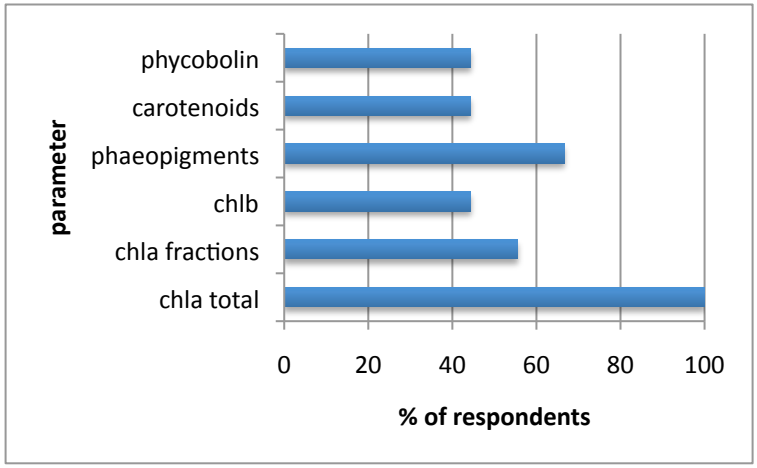

Figure 12-10: Pigments analysis evaluated by at least 40% of sampling Sites as critical or desirable for their sampling practices; chla – chorophyll a, chlb – chlorophyll b.

- **Rate and damage measurements**

For none of the responding sampling groups is evaluation of damage aspects, such as disease or infection, highly relevant.

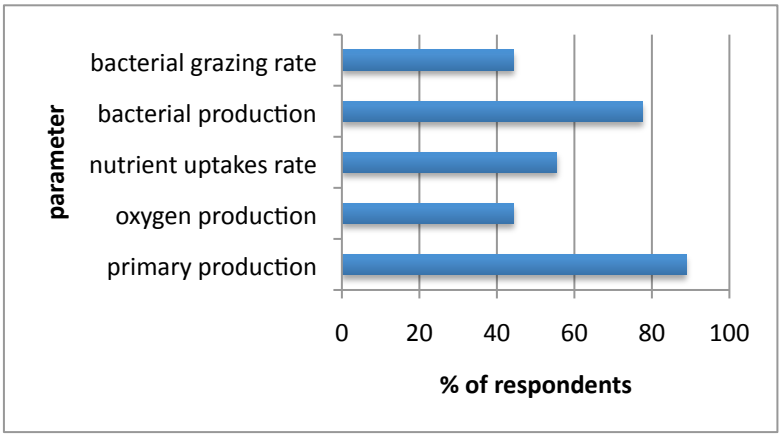

Figure 12-11: Rate measurements assessed by at least 40% of sampling Sites as critical or desirable for their sampling practices.

- **Imaging**

Images from a **fluorescent microscopy** is for 4 groups critical and for 2 groups desirable. **Electron microscopy** is for one group critical and for one desirable. None of the respondents value highly confocal microscopy or underwater photos.

- **Methodology**

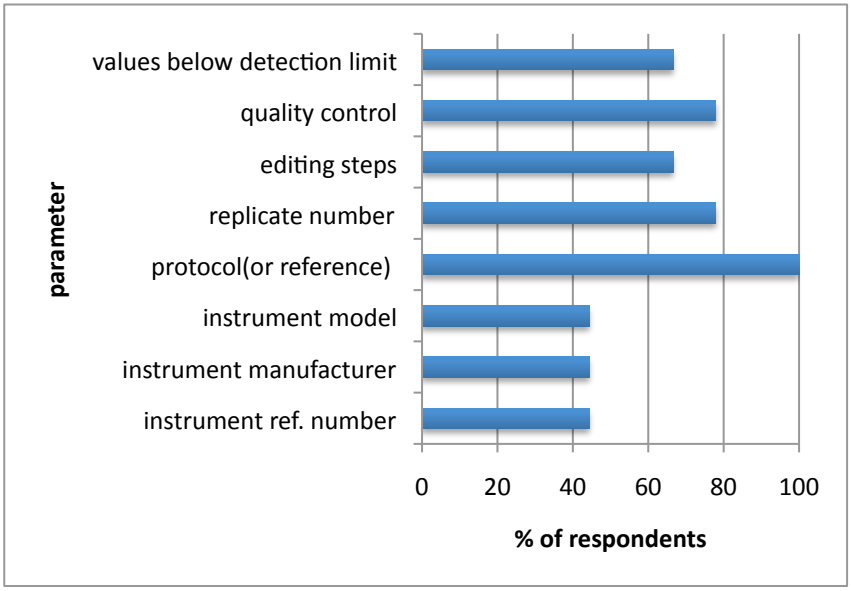

Figure 12-12: Methodological parameters assessed by at least 40% of sampling Sites as critical or desirable for interoperability of their data obtained during sampling.

- **Data archiving**

All respondents archive obtained data and metadata in their Site-specific or national database using database-specific identifiers. 5 sampling Sites have restricted access to their data at least until they are published.

Data archiving will further be addressed during development of the MicroB3 infrastructure, where oceanographic and genomic repositories will establish mutual interoperability in order to archive and provide open access to the MicroB3 genomic data in the environmental context.

Results of this Survey provide a window into the highly diverse world of sampling practices and can be indicative of the scientific parameters we shall focus on in the MicroB3 standards concepts development.
However, underlying relevant protocols for obtaining many of these parameters is a separate issue. For some groups, who sample for instance in extreme conditions, can standardization of protocols be hardly possible. For other groups can a standardized protocol, differing from their routine sampling method, be an interesting experiment and an opportunity to explore a new approach and share experiences and data with other sampling groups.
In the process of the protocols standardization we shall therefore aim for both, providing standardized set of protocols for those who like to establish new- or challenge their current methodologies, and providing the option to describe methodology as a structured text for those who prefer to use their sampling Site-tailored methodology but would like their data to be as much as possible interoperable with results of other marine microbial sampling Sites.

Annex 1

Sampling practices consensus

The overview of routine data recording presented below is not a comprehensive list of parameters captured at each sampling Site. It summarises sample data capture only during one routine sampling event reflecting consistency vs. variation of the routine sampling. This summary does not address sample processing or data processing procedures. It reviews data from six sampling Sites available at the time of the Sampling Groups Workshop in July 2012.

MIMARKS wate	**pilot OSD**	**Blanes**	**Helgoland**	**Rothera**	**Naples**	**Roscoff**	**L4**
mandatory							
project_name	project_name						
	Sample_name	Sample_name	Sample_name	Sample_name	Sample_name	Sample_name	Sample name
collection_date	collection_date	collection_date	collection_date	collection date	collection_date	collection_date	collection_date
	Description					Description	
	Title					Title	
geo_loc_name	Country					Country	
	Emp_status						
	Public					Public	
	Sample_location						
	Sample_progress						
biome	Env_biome					Env_biome	
feature	Env_feature					Env_feature	
material	Env_matter					Env_matter	
	Taxon_id					Taxon_id	
	Temp (unit)	Temp	Temp	Temp	Temp	Temp	Temp
	Salinity (unit)	Salinity	Salinity	Salinity	Salinity	Salinity	Salinity
				Density	Density		Density
	Diss_oxygen (units	Oxygen	Oxygen		Oxygen	Diss_oxygen	Oxygen
				oxyg_isotop_compos.			
	Conductivity (unit)						
	Air_saturation (unit)						humidity
lat_lon	Lat	Lat		Lat		Lat	Lat
	Lng	Lng		Lng		Lng	Lng
	Error_radius (unit)						
depth	Depth (unit)	Depth(Always surface)		Depth	Depth	Depth	Depth
		Secchi depth (m)	Secchi depth (m)		Secchi depth (m)		Secchi depth (m)
				tot_depth_water_col		tot_depth_water_col	
	Time_local	Time		Time			Time
	Time_UTC						
	Sample_volume (unit)				Sample_volume (unit)		Sample_volume (u
	Filters_stored_at (Celsius)			samp_store_temp		samp_store_temp	samp_store_temp
	Extracted_dna_avail_now						
	Physical_samp_avail_now						
	sample_id (calculated)						
	space_time_id (calculated)						
						sample_collection_device	
							wind speed
							wind direction
							rainfall
							turbidity
							aerosol_size
							aerosol_optic_thick
						altitude	
						elevation	
						assigned_from_geo	
						air temp	air temp
						atmosh. pressure	atmosh. pressure
		Light (uE m-2 s-1)		Light (uE m-2 s-1)		Light (uE m-2 s-1	Light (uE m-2 s-1)
		pH	pH	pH		pH	
		Chla (total)	Chla (total)	Chla (total)	Chla	Chla	Chla
		Chla (<3um)		Chla (20,5,2,0.2um)			
					Chlb		
					phaeophytin a		
					phaeophytin b		
		ammonium	ammonium		ammonium	ammonium	ammonium
		Nitrate	Nitrate	Nitrate	Nitrate	Nitrate	Nitrate
		Nitrite	Nitrite	Nitrite	Nitrite	Nitrite	Nitrite
					total_inorg_nitrogen		
							total_org_nitrogen
						part_org_nitrogen	part_org_nitrogen
		dissol_org_matter					col_dissol_org_mat
						suspend_part_matter	
		total_org_carbon					total_org_carbon
		part_org_carbon				part_org_carbon	part_org_carbon
		dissol_org_carbon		dissol_org_carbon			
		part_org_phosphorus					
		Phosphate	Phosphate	Phosphate	Phosphate	Phosphate	Phosphate
		Silicate	Silicate	Silicate	Silicate	Silicate	Silicate
		primary_production			prim_production (C14)		
		HPLC (total)					HPLC (pigments)
		HPLC (<3um)					HPLC (biogeochem
		Flow Cytometer					Flow Cytometer
					fract_biomass_m	organism_count	Zooplankton- abun
		Virus-DNA	Phytoplankton		fract_biomass_nano		Phytoplankton-abu
		Virus-abundance			fract_biomass_pico		Phyto-flagellates
					total diatoms		Diatoms
					Centric diatoms < 10 µm		Coccolithophores,
					Centric diatoms < 5 µm		Phaeocystis,
					Centric diatoms > 10 µm		Autotrophic dinofla
					Pennate diatoms <10 um		Heterotrophic dinof
					Pennate diatoms > 10 um		Zoo- flagellates
					total dinoflagellates		Ciliates
					Naked dinoflagellates < 15 µm		Copepod_egg_prod
					Naked dinoflagellates > 15 µm		
					Thecate dinoflagellates < 15 µm		
					Thecate dinoflagellates > 15 µm		
					Total coccolithophores		
					Undetermined phytoflagellates < 10 µm		
					Undetermined phytoflagellates > 10 µm		
					total other flagellates		
					individual species		individual species
		Bacteria-DAPI					
		FISH					
		DNA(>3um)					DNA
		DNA(3-0.2um)					
		RNA(>3um)					
		RNA(3-0.2um)					
		bacterial_production					