Marine

Microbial Biodiversity,

Bioinformatics & Biotechnology

Grant agreement n°287589

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

Deliverable 4.1

Use case workshop and e-conference report

Version: 1.0
Circulated to: Name (Date)
Approved by: Name (Date)

Expected Submission Date: 30.09.2012
Actual submission Date:

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

On 19th-20th of April 2012 the EMBL-EBI organized the MicroB3 Standards and Interoperability Use Case Workshop. Participants from six Institutes, CNRS (France), CIESM (HQ Monaco), UniHB-MARUM (Germany), JacobsUni (Germany), EMBL-EBI (UK) and UOXF (UK), analyzed and discussed prototype Use Cases in marine ecosystem biology, particularly diatom biology, and their legal interpretation in order to identify hypotheses that will be tested on the sampled marine microbial systems. In addition, Use Cases in environmental sciences and their legal implications were identified before and after the Workshop via e-communication with the MicroB3 Consortium partners.

The Use Case Workshop also crystallized the process of Use Cases transformation into a set of parameters, collectively called the MicroB3 Candidate Checklist. This list contains *i/* information elements that are specifically used or sampled for in the Use Case studies in order to answer the scientific questions the studies postulated and *ii/* information elements of minimal requirements of the Access and Benefits Sharing.

All identified Use Cases, selected from the area of diatom biology and from the area of marine prokaryotic biodiversity as well as the scientific and legal component of the Candidate Checklist are described in the related *Use Case Document*.

The Candidate Checklist will bring the hypotheses-driven input into the development of the MicroB3 standards concepts. The Use Cases will become a useful tool for testing the developed concepts of the MicroB3 standards. This standardization effort, complying with the provisions of the Nagoya protocol and enabling consistent description of marine microbial samples, shall lead to easier handling of data derived from ocean samples and optimal use of the data by the marine science community.

Table of content

Use Case Workshop and e-conference objectives

This document reports on the Use Case Workshop organized by the EMBL-EBI, Hinxon, UK, on 19th and 20th April 2012 and on e-communication preceding the Workshop.

The Use Case (UC) Workshop and the e-communication aimed to develop and discuss prototype Use Cases from marine ecosystem biology in order to identify questions addressed to and hypotheses tested on the sampled marine microbial systems.

A legal interpretation of the scientific Use Cases was an important aspect also discussed in depth at the Workshop and thereafter.

Development of the Use Cases after the Workshop and the specific role of the Use Cases in the whole MicroB3 standard and interoperability structures is outlined in the related *Use Case Document*.

The *Use Case Document* describes identified Use Cases selected from the area of diatom biology and the area of marine prokaryotic biodiversity. It also brings to light the process of Use Cases transformation into a set of scientific and legal parameters called collectively the MicroB3 Candidate Checklist. This list consists of information elements that are specifically used in the Use Case studies in order to answer the scientific questions the studies postulated, and contains a legal frame addressing traceability and fair sharing of data derived from the sampled resources. Both the scientific and the legal component of the Candidate Checklist are presented in the *Use Case Document*.

Prototype Use Cases in biological sciences and their legal interpretation were identified directly at the UC Workshop. Use Cases in environmental sciences were prepared for discussions but not addressed at the workshop itself. Information was gathered from relevant peer-reviewed scientific literature and e-communication with Prof. Jack Gilbert from Dept. of Ecology and Evolution, Univ. of Chicago, USA, and Michèle Barbier, CIESM, Monaco.

Use Case Workshop participants

The following Institutes, all MicroB3 partners, were represented at the UC Workshop

CNRS (Chris Bowler), CIESM (Michèle Barbier), MARUM (Stephane Pesant), JacobsUni (Frank Oliver Gloeckner, Julie Schnetzer), EMBL-EBI (Guy Cochrane, Stephane Riviere, Petra ten Hoopen, Charles Cook), UOXF (Peter Sterk)

Use Case Workshop agenda

19th April afternoon (EBI - Courtyard Room)

12.00 – 1.00	- working lunch at the Courtyard Room
1.00 – 1.30	- introduction to the MicroB3 Standards and Interoperability Use Case Workshop
1.30 – 3.30	- presentation and discussion of the Use Case in biological sciences
3.30 – 4.00	- coffee break
4.00 – 600	- presentation and discussion of the legal interpretation of the Use Cases
7.30	- dinner at Ickleton Lion

20th April morning (EBI - A2-33 room)

9.00 – 10.30	- identification of prototype Use Cases in biological sciences and demonstration of transformation of one selected Use Case
10.30 – 11.00	- coffee break
11.00 – 12.00	- summary and wrap up
12.00	- working lunch

Selected scientific domains

Three scientific studies have been selected prior to the UC Workshop and presented to the Workshop participants for detailed discussions.

The next section introduces the studies, explains reasons for their selection and summarizes methods used in the studies.

Study in biological sciences

Marine microbial communities are incredible diverse, consisting of cyanobacteria, heterotrophic bacteria, archaea, viruses, eukaryotic phytoplankton and protists.

As a prototype for the use case in biological sciences we have chosen diatoms - eukaryotic phytoplankton that evolved from endosymbiosis of red algae, green algae and eukaryotic heterotrophs. This most diverse group of phytoplankton combines animal- and plant-like abilities and contains in the genome bacterial and viral genes acquired by horizontal gene transfer.

One-fifth of the photosynthesis on Earth is carried out by diatoms, which are considered a buffer of climate changes. They dominate in coastal and upwelling regions, are a critical component of the food chain in polar regions and serve as a biological carbon pump due to their mobility through the sea water column.

The study Bowler *et al*., 2010 highlights the dependence of diatom cell cycle progression on environmental factors. Particularly the G1 and G2 phase of the cell cycle are critical checkpoints dependent on light (intensity and wavelength), nutrients (iron, nitrate, phosphate). The G2 phase is in many diatom species dependent on the availability of silicate.

Seasonal variability in wind-driven sea currents leads to changes of temperature and salt concentrations. Upwellings draw nutrient- and CO_2-rich water from lower layers to the surface and together with light penetrating to the water column feed diatom blooms.

The above study suggests that a number of parameters should be recorded at any given time and any given region in order to investigate diatom-specific processes.

Discussions were expected to focus on factors that can influence seasonal changes of phytoplankton blooms, their proliferation and termination and shall therefore be in standardized and consistent ways recorded during any sampling event.

*Methods used in *Diatom cell division in an environmental context*:
Authors have used comparative and functional genomic analyses to provide molecular information about diatom cell division mechanisms, and their regulation in response to environmental conditions. Three approaches have been used: quantitative PCR and cDNA-AFLP to examine gene expression in synchronized cells, and timelapse imaging of fluorescently labelled cellular structures during cell division.

Study 1 in environmental sciences

We have considered the exploration of the microbial seed bank in the Western English Channel (WCO) L4 marine observatory a very suitable prototype Use Case in the environmental sciences. A deep study of the WCO seasonal changes in microbial community and of the WCO microbial metabolome provide a significant contribution to the understanding of global ecology.

The study Caporaso *et al.*, 2012 showed that seasonal dynamics in bacterioplankton diversity can be explained by modulation of relative abundance of taxa that are always present in the L4 site and it is not a result of taxa extinction and recolonization in the ecosystem through time.

The authors have compared data of one deep Illumina sequencing of the hypervariable V6 region of the 16S rRNA gene at a single time point with shallow 454 sequencing results taken at 72 time points spanning a period of 6 years and concluded that seasonal differences in marine microbial community do not include changes in community membership.

Discussions was expected to focus on factors influencing the structure of populations, which should be considered and recorded if a similar study took place at various geographic coordinates, various time points or water depths.

*Methods used in *The Western English Channel contains a persistent microbial seed bank*:
Monthly samples subject to amplicon pyrosequencing of the V6 hypervariable region (68 base pairs) of the 16S rRNA gene. Illumina sequencing performed using standard library prep reagents, and v2 clustering and v3 SBS kits. The Illumina paired-end reads were filtered to remove any sequences containing adaptors and then quality filtered during paired end merging, by requiring that at least 10 bases on each end overlapped, and overlapping bases exactly identical.

Study 2 in environmental sciences

The study Larsen *et al.*, 2011 applied a method of Predicted Relative Metabolic Turnover (PRMT), which infers metabolic activity from unique enzyme functions encoded in the metagenome, to predict changes in turnover of metabolites in a marine metabolome of the Western Channel Observatory.

Since it is not easily possible to measure every metabolite in an environment, a careful determination of pertinent environmental parameters, that will allow formulate a testable biological hypothesis, is an essential prerequisite.

Correlation of measured environmental parameters with predicted changes in turnover of metabolites and changes in relative abundance of bacterial phyla generated a number of interesting hypotheses, such as linking measured environmental concentrations of ammonia to bacterial catabolism of chitin and identifying Actinomycetes as potent chitin degrading bacteria.

Discussions were expected to identify environmental parameters that should be monitored by ocean sampling projects to allow better data comparison and investigation of biological hypotheses, e.g. by using the PRMT analysis?

*Methods used in Predicted Relative Metabolomic Turnover (PRMT): determining metabolic turnover from a coastal marine metagenomic dataset:
Predicted Relative Metabolic Turnover (PRMT) computational method based on Environmental Metabolome Matrix and normalized Enzyme Activity Counts.

First day

Guy Cochrane, leader of the MicroB3 – Work Package 4, welcomed all to the Use Case Workshop and invited participants of the meeting to the "tour de partners", where everybody shortly introduced himself.

Guy Cochrane (EBML-EBI) then gave a brief introductory talk to the MicroB3 Work Package 4 (Figure 1), Stephane Riviere (EBML-EBI) informed about the Workshop report as the first WP4 deliverable to the MicroB3 project and Petra ten Hoopen (EBML-EBI) presented details of the Workshop agenda.

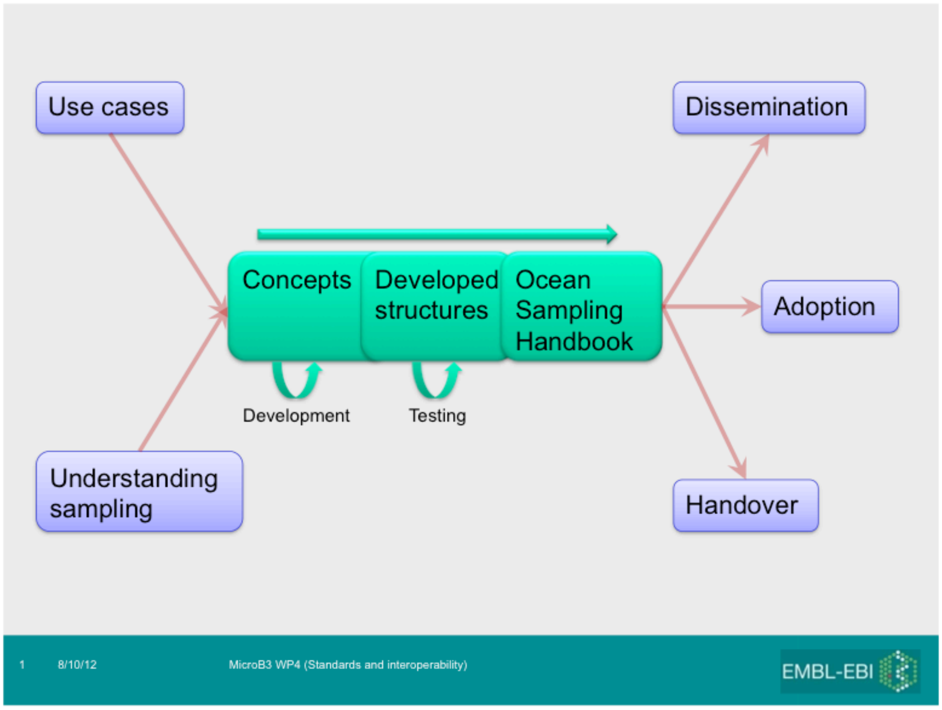

Figure 1: Overview of the MicroB3 – Work Package 4 interoperability structures development.

In the next session Chris Bowler (CNRS) gave a review talk (Figure 2) on diatom biology – the main focus of the biological prototype Use Cases.

This comprehensive overview informed discussions that started to shape scientific hypotheses relevant to diatom biology. Discussions related to

- monitoring of diatom diversity in nutrient-rich upwelling zones dominated with diatomes during blooms
- dependance of diatoms on silicate
- diatoms periodic proliferation and termination and diatoms as major contributors to the biological carbon pump
- endogenous signals and environmental factors controling diatom life cycle

- disproportion between tremendous genetic diversity of diatoms and minimal available nucleotide sequence information

As a byproduct of these discussions a number of issues arisen that relate to practical aspects of sampling

- which standard device should be present at each sampling station
- how will samples be distributed – in aliquots or will there be one appointed lab
- how to provide integrated access to the acquired data

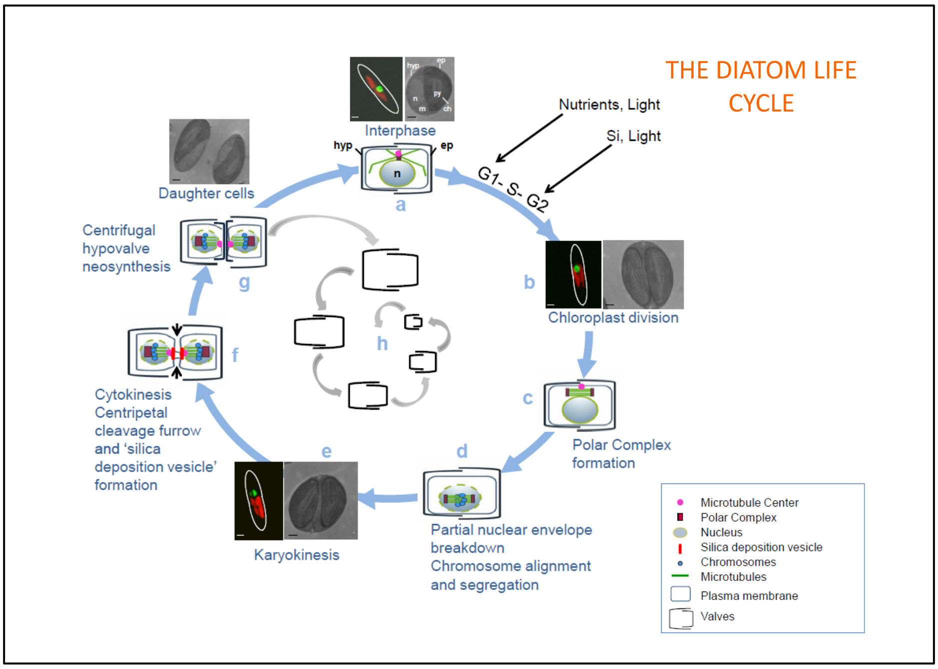

Figure 2: Introduction to the review on diatom biology (upper) and illustration of the diatom life cycle (lower, Bowler *et al.*, 2010).

In the following session Michèle Barbier (CIESM) presented discussion points to the minimal requirements for the fair and equitable sharing of benefits (Figure 3) derived from the access and use of genetic resources (Access and Benefits Sharing, ABS). Her talk also initiated discussion on a need to define The Code of Conduct for the OSD participants. The Mediterranean Code of Conduct is a morally binding agreement (innitiated and developed at CIESM, within the WP8 of the MicroB3 project) supporting the ABS during the Ocean Sampling Day in the Mediterranean.

- Provide evidence of origin of a resource (including the geographical origin of resources and of compliance with the access laws of the providing country)
- Identify the source that cover MGR provided by primary resources and secondary resources (ex situ collections, databases on genetic resources and scientific literature)
- Track flows/monitoring of the utilisation of genetic resources
- Identify national permitting procedures facilitating implementation of user measures and checkpoints. Provide legal certainty regarding rights to use resources
- Define PIC, MAT and/or fair and equitable benefit sharing (how it can support indigenous people and local communities resources and management practices)
- Demonstrate compliance with domestic ABS and/or TK legislation
- Identify a market use of resources and TK (e.g. biopharmaceuticals, cosmetics, processed food)

Figure 3: Minimum requirements for ABS (CIESM proposal).

A number of legal issues were discussed

- how to make sure generated data are properly credited and cited
- how we share resources from the open ocean where the Nagoya Protocol does not apply
- where stand genomic and oceanographic repositories, that will archive majority of the OSD data, in terms of Nagoya Protocol
- whether there should be a period of moratorium on the OSD data that will be archived in public databases

Second day

Participants of the UC Workshop formulated four diatom-related biological Use Cases (Figure 4).

area	scientific aspect	legal aspect
1. Diatom diversity as a function of latitude and seasonality	Characterize diatom taxonomic profiles using 18S, images, biomass	Legal development of fair and equitable use of resources through assessment of biodiversity according to national economic status
2. Diatom silica biochemistry	Characterize silicate metabolism and cell wall formation and their regulation	Industrial exploitation of silicate metabolizing pathways
3. Diatom carbon pump	Assess environmental influences upon carbon pump activity	Ecosystem services – legality and ethics of intervention
4. Regulation of sexual/asexual life cycle	Understanding environmental triggers and regulators of switching between strategies	Exploitation of diatoms as ocean health bio-indicators for global monitoring programs

Figure 4: Prototype Use Cases identified in the field of diatom biology.

In the second session participants worked together on a transformation of one selected Use Case – the Use Case 4 - Regulation of sexual/asexual diatom life cycle (Figure 5).

The first set of scientific and legal parameters have been identified that should be captured in order to be able to answer the scientific question of this Use Case and comply to the ABS.

This list of information components expanded when transformation of other Use Cases has been performed. The Candidate Checklist, described, as mentioned above, in the *Use Case Document*, sums parameters identified in all prototype Use Cases.

The Use Case Workshop has been summarized with suggestions of next steps for further Use Cases identification and analysis.

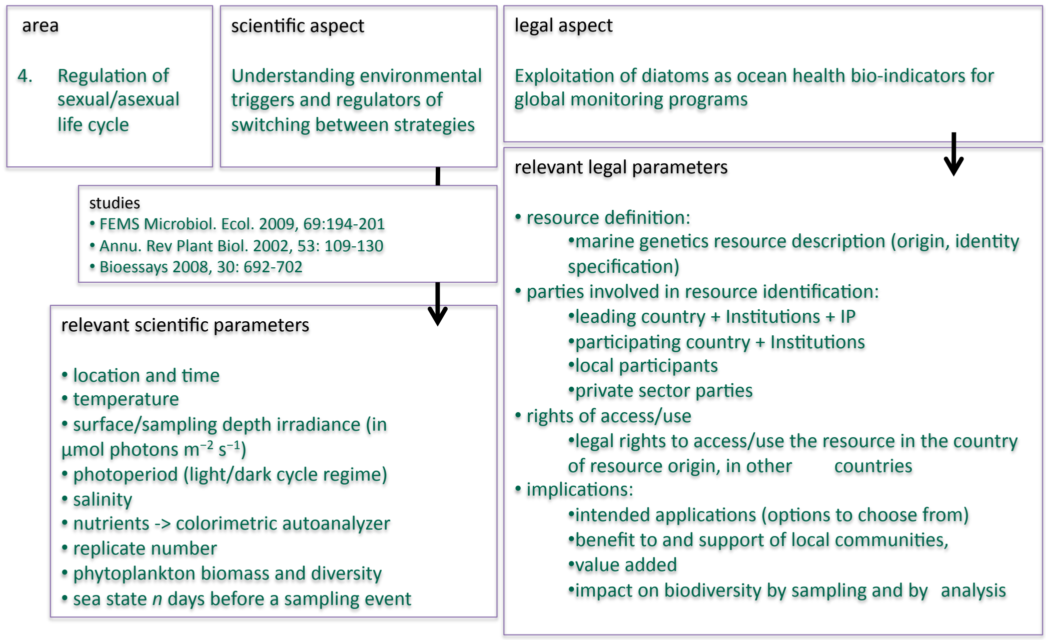

Figure 5: Example of a Use Case transformation into a list of information components needed to answer questions the Use Case addresses and to comply to ABS.

Based on these discussions the following steps were taken to further develop the prototype Use Cases and to formulate the Candidate Checklist.

- transformation of Use Cases in biological sciences based on relevant literature
- transformation of the Use Cases in environmental sciences, based on relevant literature and e-communication with authors of the studies 1 and 2 in environmental sciences
- draft legal parameters relevant to the identified Use Cases in both biological and environmental sciences
- develop a draft of the Candidate Checklist of scientific and legal parameters
- obtain a feedback on the draft of the Candidate Checklist from the participants of the UC Workshop
- optimize the Candidate Checklist based on the feedback
- present the Candidate Checklist to the participants of the Sampling Groups Workshop in summer 2012

Reference list

Bowler Ch, De Martino A, Falciatore A (2010) Diatom cell division in an environmental context. Current Opinion in Plant Biology 13:623-630

Caporaso JG, Paszkiewicz K, Field D, Knight R, Gilbert JA (2012) The Western English Channel contains a persistent microbial seed bank. ISME Journal 6:1089-1093

Larsen PE, Collart FR, Field D, Meyer F, Keegan KP, Henry CS, McGrath J, Quinn J, and Gilbert JA (2011) Predicted Relative Metabolomic Turnover (PRMT): determining metabolic turnover from a coastal marine metagenomic dataset. Microbial Informatics and Experimentation 1:4

The Use Case Document

MicroB3 Standards and Interoperability – Work Package 4
September 2012

Petra ten Hoopen, Guy Cochrane
EMBL-EBI, UK

Acknowledgement
We would like to acknowledge Stéphane Pesant (MARUM, Germany), Michèle Barbier (CIESM) and Chris Bowler (CNRS, France) for their valuable input into the prototype use cases development.

Summary

The MicroB3 Standards is a collection of parameters and controlled vocabularies enabling consistent description of marine microbial samples and sampling methods, formulated in compliance with the provisions of the Nagoya protocol, and allowing better handling of data derived from the ocean samples for their optimal use by the MicroB3 consortium partners as well as by the broader marine science community.

This document outlines the process of the MicroB3 Standards development and particularly focuses on the development of the MicroB3 prototype Use Cases.

We describe here identification of scientific and legal aspects of the prototype Use Cases and their transformation into the scientific and legal component of the Candidate Checklist.

The MicroB3 prototype Use Cases were selected from the area of diatom biology and from the area of marine prokaryotic biodiversity based on information gathered at the MicroB3 Use Case Workshop, from e-discussions with the MicroB3 Consortium partners and relevant peer-reviewed scientific evidence.

These prototype Use Cases will become a tool for testing of MicroB3 Standards Concepts. Transformation of the prototype Use Cases in the form of the Candidate MicroB3 Checklist will become the hypotheses-based input into the development of the MicroB3 Standards.

Table of contents

MicroB3 Work Package 4 objectives and strategy

The main objective of the MicroB3 project is to create a platform for obtaining, smooth handling and analysis of marine viral, bacterial, archaeal and protists data in an environmental context. Interdisciplinary links will be established between biologist building genome collections and databases archiving genomes/metagenome sequences on one side and experts in oceanography and databases storing contextual data on marine ecosystem biology and modeling on the other side.

MicroB3 standards and interoperability structures should therefore reflect the link between molecular microbial research and oceanography.

The MicroB3 Work Package 4 (MicroB3 - 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.

The developed standards shall

- provide a balance between richness and compliance of marine data
- respect established procedures
- encourage rich and consistent description of a sampling site and sample processing (as a collection of parameters containing scales, structured text and controlled vocabularies)
- include minimal reporting requirements, i.e. a core list of parameters to be reported, and standard operating procedures
- cover generation, analysis and archiving of data derived from the ocean samples
- consider legal aspects of marine sample collection

The interoperability structures should support dissemination of the data to the scientific community for a long-term utility and facilitate implementation of the integrated view of marine microbial diversity.

Several streams of information will feed into the MicroB3 concepts development (Figure 1).

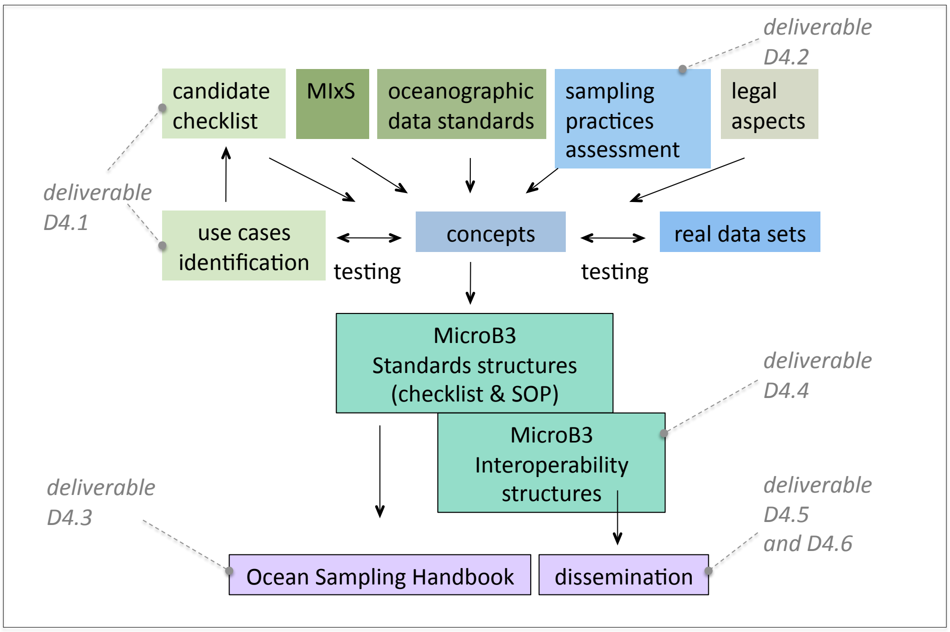

Figure 1: Schematic representation of information streams providing input into the MicroB3 standards concepts development and output from the finalized structures.

The first, hypotheses-based source of information originates from identification of suitable prototype Use Cases and their transformation into the minimal candidate checklist. This has been addressed at the Use Case Workshop, organized by the MicroB3 - WP4 in April 2012, and followed up in e-discussions with the MicroB3 Consortium partners. Results are reported in *the MicroB3 deliverable D4.1*.

The second source of information comes from understanding current- and establishing best sampling practices. It was the main task of the Sampling Groups (SG) Workshop, organized by the MicroB3 - WP4 in July 2012, and the aim of the Sampling Groups Survey, which reviews information components currently being captured at the MicroB3- and other external Sampling Sites. Outcomes of the SG Workshop and analysis of results of the SG Survey are reported in *the MicroB3 deliverable D4.2*.

Concepts development will inevitably draw information from mapping the candidate checklist to genomic MIxS and oceanographic standards currently in place.

The overall aim of the MicroB3 project is to provide genomic information in both environmental and legal context. The minimal checklist will therefore include legal parameters assuring traceability across the whole supply chain, from geographic source to the end use and marketing of the material.

Information components will be evaluated and critical, desirable and useful aspects of a sampling site, a sample processing and a data handling will be identified using counterfactual techniques applied to Use Cases.

The developing concepts, that provide a balance between optimal and feasible, will be tested using concept-compliant real data sets.

The MicroB3 standards and interoperability concepts will eventually evolve into minimal reporting requirements and Standard Operating Procedures (SOP) that will be published in the Ocean Sampling Handbook for use by the Ocean Sampling Day participants and disseminated to the broader marine science community.

MicroB3 standards in context of other standardization efforts

The Genomic Standards Consortium (GSC, http://gensc.org/gc_wiki/index.php/Main_Page) provides the microbial community with minimum information checklists for genomic data (MIGS), metagenomic data (MIMS) and environmental sequences including ribosomal genes (MIMARKS). These standards, collectively called MIxS, were designed to describe specific information about genomic sequences in order to provide harmonious capture, collection and exchange of metadata. All MIxS standards contain shared core descriptors, checklist-specific descriptors and each can be extended by an optional environmental packages, specific for a certain environment, such as air, a water column or sediment.

Similarly, a number of initiatives aim to standardize oceanographic data, only a few of which can be mentioned here.
The EnvEeurope (http://www.enveurope.eu/) project objective is a harmonization of assessment methodologies for selected environmental parameters to reach compatibility of data gained at the long-term ecosystem research monitoring (LTER) sites across Europe.
Marine environmental data & information network (MEDIN, http://www.oceannet.org/) provides a set of common standards for metadata, data format and content agreed upon and supported by MEDIN partners.
The new SeaDataNet Common Data Index profile of the ISO 19115 metadata content standard (http://www.seadatanet.org/content/download/13457/89112/file/SDN-ISO19115-profile-V7.0.pdf) can be considered a standard for marine metadata in Europe.
Consortia exploring global ocean biodiversity, such as TARA Ocean (http://www.plosbiology.org/article/info%3Adoi%2F10.1371%2Fjournal.pbio.1001177) or projects such as GOS (Global Ocean Sampling, http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0050077) developed sampling protocols allowing respectively holistic or global genomics studies of marine biodiversity.

Expertise of these genomic and oceanographic standardization efforts will have a valuable input into the development of the MicroB3 standards that should provide a common conceptual ground broad enough to cover different aspects of marine microbial ecosystems.

The MicroB3 prototype Use Cases

A prototype Use Case for our purposes is a scientific question/hypothesis in biological or environmental sciences that will be asked of the sampled marine microbial systems.

Specific, testable hypotheses shape sampling strategies. Identification and analysis of suitable prototype Use Cases provides therefore important hypotheses-driven contribution to the standards and interoperability structures development.

We have identified with the assistance of the MicroB3 Consortium partners and stakeholders four biological and two environmental prototype Use Cases.
The biological Use Cases focus on diatom biology. The environmental Use Cases relate to the marine prokaryotic biodiversity.

The *MicroB3 deliverable D4.1* outlines in the section *Selected scientific domains* reasons for selecting Use Cases from these particular areas of marine microbial ecosystem biology. Therefore, we will not duplicate this piece of information here and refer an interested reader to the above mentioned section.

All prototype Use Cases, their scientific aspect and legal interpretation are summarized below.

A. Prototype Use Cases in Biological Sciences

1. Diatom diversity as a function of latitude and seasonality

Scientific - Characterize diatom taxonomic profiles using 18S rRNA gene, images, biomass.
Legal – Legal development of fair and equitable use of resources through assessment of biodiversity according to national economic status

2. Diatom silica biochemistry

Scientific – Characterize silicate metabolism and cell wall formation and their regulation
Legal – Industrial exploitation of silicate metabolizing pathways

3. Diatom carbon pump

Scientific - Assess environmental influences upon carbon pump activity
Legal – Ecosystem services – legality and ethics of intervention

4. Regulation of sexual/asexual life cycle

Scientific – Understanding environmental triggers and regulators of switching between strategies
Legal – Exploitation of diatoms as ocean health bio-indicators for global monitoring programs

B. Prototype Use Cases in Environmental Sciences

1. Seasonal changes in microbial community composition at multiple specific locations

Scientific – Characterize seasonal and spatial changes in microbial pattern using 16S rRNA gene
Legal – Adaptation capability of marine microbial community to changing environments caused by natural and anthropological factors

2. Predictions of microbial community structures

Scientific – Modeling and prediction of microbial life in time and space
Legal – Use of community structure predictions to mitigate negative consequences of environmental changes

The MicroB3 candidate checklist

The MicroB3 Candidate Checklist has two components, scientific and legal.

The **scientific component** of the MicroB3 Candidate Checklist is defined as a list of parameters, i.e. measurable variables, we would optimally need to capture to answer the scientific questions postulated in the prototype Use Cases.

The **legal component** is an additional layer comprising parameters ensuring that requirements of traceability, legality and ethics across the whole chain, from the sample collection to the end use of the acquired data, are satisfied. This non-exhaustive list of legal information elements will further be developed in collaboration with the MicroB3 Work Package 8.

The MicroB3 Candidate Checklist was developed in a three-step process:

Step 1.
A counterfactual technique was applied and intersection of the prototype Use Cases performed using peer-reviewed publications relevant to each Use Case area.
As a result of this transformation information elements were identified that are specifically used in the diatom studies or specifically used in the prokaryotic biodiversity studies. These selected fields we would optimally need to capture during marine sampling in order to test hypotheses postulated in the prototype MicroB3 Use Cases.

Step 2.
The selected fields were consulted with experts from MARUM (Stéphane Pesant) and CNRS (Chris Bowler) in biological sciences, and experts from Univ. of Chicago, USA (Jack Gilbert) in environmental sciences.
Based on these e-discussions the scientific component of the Candidate Checklist has further been optimized and presented to the participants of the Sampling Group Workshop in July 2012 (Table 1).

Step 3.
Legal interpretation of each Use Case has been considered and legal aspects transformed into a candidate list of legal parameters based on consultation with experts from the CIESM (Michèle Barbier).

The legal parameters should secure realization of a key objective of the Convention on Biological Diversity, which is the fair and equitable sharing of benefits derived from the access and the use of genetic resources. It should be a system for traceability across the whole supply chain, from geographic resource to end use and marketing of the material.

The **scientific component** of the MicroB3 Candidate Checklist is available in the **Annex 1**. It will become the hypothesis-driven input into the MicroB3 standards structures development.

The **legal component** of the MicroB3 Candidate Checklist is available in the **Annex 2**. It is a non-exhaustive list of information components the scientists will need to consider and provide before any sampling event.

For the convenience of the scientists Michele defined a *Project*, which groups the information elements included in the legal component of the MicroB3 Candidate Checklist required to define PIC (Prior Informed Consent) and MAT (Mutually Agreed Terms) for obtaining access to the sampled Material. This *Project* definition, with no legal value, refers to one scientific expedition including material sampling, its transfer, parties involved and use of the material.

Project

- project name
- project description (objectives and purposes)
- provider – the legally entitled provider/ national competent authority
- recipient - main PI and Institution (name, address, contact)
- third parties details (scientists from research institutes, scientists from private companies, scientists from local research institutes, scientists from local private companies) participating during
    - collection/sampling
    - transfer
    - analysis
    - use
    - storage
- the Material,
    - name and description
    - exact location (country, GPS, etc.)
- the expedition description
    - vessels name/ cruise identifier
    - country / flag
    - technical description
    - duration of the expedition
- equipment/technologies and methods used for
    - collection/sampling
    - transfer

- analysis
- storage

- physical location of the material during
  - transfer
  - analysis
  - storage/biorepositories

- Third parties implicated during
  - collection/sampling
  - transfer
  - analysis
  - storage

- the use of the Material **-** description of the potential utilization of the material (commercial/non-commercial purposes):
  - which use of the material and results (publication, ex situ collection, commercial application)
  - identifiable end product, taking into consideration other potential types of use or derivatives and products of the resources (e.g. research, breeding, commercialisation),
  - description of benefits of sharing
  - description of property rights

- infrastructure storing the data on the long term
- infrastructure storing the resources on the long term

The identified Use Cases will be employed for testing of concepts of the MicroB3 standards that should provide an appropriate scale and relevant resolution for a sampling site description and methods elaboration beneficial to all parties involved.

References

The **Prototype Use Cases in Biological Sciences** were analyzed based on the following studies

Bowler C, De Martino A, Falciatore A (2010) Diatom cell division in an environmental context. Current Opinion in Plant Biology 13:623-630 (link)

Mouget JL, Gastineau R, Davidovich O, Gaudin P, Davidovich NA (2009) Light is a key factor in triggering sexual reproduction in the pennate diatom *Haslea ostrearia*. FEMS Microbiology Ecology 69:194-201 (link)

Falciatore A, Bowler C (2002) Revealing the molecular secrets of marine diatoms. Annu. Rev Plant Biol. 53: 109-130 (link)

Olson RJ, Vaulot D, Chrisholm SW (1986) Effects of environmental stresses on the cell cycle of two marine phytoplankton species. Plant Physiology 80: 918-925 (link)

Ragni M, D'Alcala MR (2004) Light as an information carrier underwater. Journal of Plankton Research 26:433-443 (link)

Chepurnov VA, Mann DG, von Dassow P, Vanormelingen P, Gillard J, Inze D, Sabbe K, Vyverman W (2008) In search of new tractable diatoms for experimental biology. BioEssays 30: 692-702 (link)

The **Prototype Use Cases in Environmental Sciences** were analyzed based on the following studies

Caporaso JG, Paszkiewicz K, Field D, Knight R, Gilbert JA (2012) The Western English Channel contains a persistent microbial seed bank. ISME Journal 6:1089-1093 (link)

Larsen PE, Collart FR, Field D, Meyer F, Keegan KP, Henry CS, McGrath J, Quinn J, and Gilbert JA (2011) Predicted Relative Metabolomic Turnover (PRMT): determining metabolic turnover from a coastal marine metagenomic dataset. Microbial Informatics and Experimentation 1:4 (link)

Gilbert JA, Steele JA, Caporaso JG, Steinbrueck L, Reeder J, Temperton B, Huse S, McHardy AC, Knight R, Joint I, Somerfield P, Fuhrman JA, Field D (2012) Defining seasonal marine microbial community dynamics. ISME Journal 6: 298-308 (link)

Gilbert JA, Field D, Swift P, Thomas S, Cummings D, et al. (2010) The taxonomic and functional diversity of microbes at a temperate coastal site: A ‘multi-omic’ study of seasonal and diel temporal variation. PLoS ONE 5(11): e15545. doi:10.1371/journal.pone.0015545 (link)

Larsen PE, Field D, Gilbert JA (2012) Predicting bacterial community assemblages using an artificial neural network approach. Nature methods 9:621-625 (link)

The **Legal Interpretation** was developed based on the following studies

United Nations (2011) The Nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the Convention on Biological Diversity. (link)

UNU-IAS (2008) Certificates of clarity or confusion, the search for a practical, feasible and cost effective system for certifying compliance with PIC and MAT. (link)

DFG (2008) Supplementary instructions for funding proposals concerning research projects within the scope of the Convention on Biological Diversity (CBD) (link)

Biological Resources Access Agreement between Commonwealth of Australia and J. Craig Venter Institute (2004) (link)

Annex 1

The scientific component of the MicroB3 Candidate Checklist.

1. column – parameters numbering
2. column – X marks parameters relevant to the Use Case studies in biological sciences
3. column – X marks parameters relevant to the Use Case studies in environmental sciences
4. column – parameter's name
5. – 8. column –source of information colour-coded according to preference of their provenance

	Diatom biology	Prokaryotic microb. diversity		Preferred source (green) Second preferred source (yellow) Third preferred source (red) Not applicable (grey)	Source of information: In Situ Observations or Sample collection AT the location/time of sampling/experiment	Source of information: Ancillary Observations NEAR the location/day of sampling and/or recent history (e.g. meteo station, fixed moorings, gliders/floats)	Source of information: Predictive models NEAR the location/day of sampling and/or recent history (e.g. satellite SST, SSH, Colour)	Source of information: Climatologies Statistics at various temporal & spatial scales on the location (e.g. monthly mean, min, max nutrient concentrations based on historical observations)
1	X	X		Date/Time	• GPS			
2	X	X		Latitude and Longitude	• GPS			
3	X	X		Sampling depth	• CTD or cable out (m)			
4		X		Sampling design Eulerian (a fixed site) or Lagrangian (moving with the water flow)		Cruise report or logbook		
5	X	X		Temperature	• CTD (vertical profile)	ARGO profiles	SST	Based on time series obs.
6	X	X		Salinity	• CTD (vertical profile) • Water samples	ARGO profiles		Based on time series obs.
7		X		Density	• CTD (vertical profile)			

8	X	X		Irradiance	• Underwater PAR sensor (vertical profile) • PAR sensor in air	Weather buoys or nearby land-based weather stations	Theoretical/potential	
9	X	X		Photoperiod (light/dark cycle)	•	Irradiance sensor	Theoretical/potential	Based on time series obs.
10	X	X		Wind speed and direction	• Ship's weather station	Weather buoys or nearby land-based weather stations	Wind fields from weather services, e.g. NOAA National Climatic Data Center	Based on time series obs.
11	X	X		Sea state (sea level, tidal stage)	Visual		Sea state from weather services	
12	X	X		Currents	• Ship's ADCP Lowered ADCP on a rosette	Moorings with current meters at various depth, but rarely near surface	Surface geostrophic currents from SSH	
13	X	X		Pigments concentrations	• Filtrations (chla extraction for fluorometry) • Filtrations (HPLC) • Fluorescence profile	ARGO profilers	Ocean colour	Based on time series obs.
14	X	X		Primary production	• Water sample (isotopes or probes) • In situ (isotopes or probes or chlorophyll fluorescence based measurement)		Remote sensing products	
15	X			Microplankton identification, abundance and biomass	• Water sample (Lugol) • Water sample (formol) • Water sample (Flowcam) • Water sample (digested) • Water sample (FCM) • Water sample (HTM)		Remote sensing products	Based on time series obs.
16		X		Meso- and Macrozooplankton identification, abundance and biomass	• Net samples (Vertical, oblique or horiz. tows) • Microscopic analysis • Semi-automatic imaging analysis			Based on time series obs.
17	X	X		Nutrients (phosphate, nitrate, nitrite, silicate, ammonia)	• Water samples (-20°C) • Optical sensor (needs calibration)			Based on time series obs.

18	X	X		Total, Particulate and Dissolved Organic Carbon and Nitrogen (TOC, POC, DOC) and particulate Silica	• Filtrations • Water samples • non-dispersive infrared analysis (NDIR)			
19	X			Carbonate chemistry	• Water samples pH sensor			
20	X	X		Dissolved oxygen concentration	• Water samples (titration) • O2 sensor (vertical profile)			
21		X		Soluble Reactive Phosphorus	• spectrophotometrically			
22		X		Climatic indexes			e.g. North Atlantic Oscillation (NAO) index	Based on time series obs.
23		X		Rainfall	Ship's weather station	Weather buoys or nearby land-based weather		
24	X			Organism structures and cellular activity	Water samples for imaging with fluorescent, electron, and confocal microscopy			

Annex 2

The legal component of the MicroB3 Candidate Checklist

Tentative definition of terms

Material refers to marine genetic resources collected, transferred, analyzed and stored, (microbes, protists, bacteria, viruses, etc., genes, extract, active compound, all collections of a specified species, individual samples, ex situ collections)
Transfer of Material refers to the material sampling and its transfer for analysis (nucleic acid extraction/sequencing)
Data refers to all information arising from the analysis of the material
Provider refers to the country giving the authorization to sample under its national jurisdiction
Recipient refers to the institution asking the authorization to sample under national jurisdiction
Third parties refers to entity implicated in the material collection, transfer and analysis of the material and data. (public bodies from governments, research institutes, and/or private companies and multinationals)

Parties description
(persons/institution/country/address implicated in OSD)

The Recipient
- Country
- name of the leading principal investigator
- Institution detail (address and mail)

The Provider (country giving the authorization to sample under its national jurisdiction)
- Country
- name of the National Competent Authority or legally entitled provider
- name of the collaborating scientist (institution, address etc.)

Third parties (public bodies from governments, research institutes, and/or private companies and multinationals, local research institutions or private companies)

- parties associated to the sampling/collection,
- parties associated to the transfer and analysis (nucleic acid extraction, sequencing)
- parties associated to the storage of the material/biorepositories
- parties associated to data analysis (annotation)
- parties associated to data storage

Material description (marine genetic resources)

Material
- material name
- material reference number (for traceability)
- material description using
  - taxonomic coverage (groups of living organisms sampled)
  - marine biome ontology terms at

http://www.ontobee.org/browser/rdf.php?o=ENVO&iri=http://purl.obolibrary.org/obo/ENVO_00000447
  - marine feature ontology terms at

http://www.ontobee.org/browser/rdf.php?o=ENVO&iri=http://purl.obolibrary.org/obo/ENVO_01000031
  - environmental matter ontology terms for water /sediment at

http://www.ontobee.org/browser/rdf.php?o=ENVO&iri=http://purl.obolibrary.org/obo/ENVO_00010483

Evidence of legal provenance of the Material
- date
- time local and universal
- in situ material Latitude, Longitude
- in situ water column/sediment (to be specified)
- in situ sampling depth
- name of the sampling station if relevant
- sampling duration
- country (use control vocabulary at http://insdc.org/country.html)
- region (e.g. Bay, sub-regional Sea)
- Sea/Ocean

Transfer of the Material
- physical location during analysis (nucleic acid extraction, sequencing)
- PI responsible for analysis
- Third parties implicated
- physical location during storage/maintenance

The expedition description

- vessels name/ cruise identifier
- Country / flag
- duration of the expedition

Regulations

(by structures and laws governing access to material collected under national jurisdictions)

access and use to the Material

- name of the national competent authority in the providing country
- PIC (Prior Informed consent)
- and MAT (Mutually Agreed Term)
- permit of access of the resources collected under national jurisdiction and reference number of the agreement signed to access and use of the material
- agreement on benefits sharing between parties in kind (to be specified) or monetary (if applicable) including clause on downstream commercial use of the material
- clear intellectual properties rights assignation

Biorepositories

- policies/regulation of biorepositories infrastructure (where the material is stored)
- policies regarding access to resources

Data

-OSD data shall be publically available as soon as the quality validation has been completed.
-OSD data shall be stored and accessible via three basic infrastructures, SeaDataNet, EurOBIS and ENA.
