



Protocol of the 1st internal workshop

Date:

2014 June 27th

Location:

Staatliche Naturwissenschaftliche Sammlung Bayerns, IT-Center, Menzinger Straße
67, 80796 München

Participants:

- Klaster Sabrina, University of Bayreuth
- Leininger Iris, SNSB IT Center, München
- Rambold Gerhard, University of Bayreuth
- Triebel Dagmar, SNSB IT Center, München
- Yilmaz Pelin, Jacobs University, Bremen

ITEMS ON THE MOD-CO AGENDA

1. Introduction and aim of the project
 2. Short reports of the work groups
 3. Demos of the available tools (MOD-CO Wiki, SILVA/megx.net, DiversityCollection, DiversityDescriptions) and discussion about the aspirated MOD-CO aims and tools
 4. To dos and distributions of tasks
-

1. INTRODUCTION AND AIM OF THE PROJECT

- Short summary by D. Triebel about the idea, the promise and the coming off of the DFG-supported project **Meta-omics Data Collection Objects** (abbr. MOD-CO)
 - Huge numbers of collected environmental samples exist. There are problems with storage and conservation.
 - DNA extracts, actually particularly of microorganism (maritime and terrestrial (micro-) organisms are present in many laboratories. It is therefore necessary to find ways how to curate these and also guarantee that links between of material, (sequence) data metadata are established and maintained in the long term.
 - The publication of a short press release about MOD-CO is to be found on (http://www.mod-co.net/w/media/archive/a/aa/20140603160339!Presse-MOD-CO_SNSB.pdf)
- It is the aim of MOD-CO to build the bridge between the two domains of environmental sample collections and (meta-)omics application techniques and products; therefore adequate vocabularies, schemas, models and implementations have be established:
 - Vocabularies with descriptors, definitions and examples. The vocabulary (or namespace) should follow strict rules concerning wording, style, and referencing to existing elements in other namespaces.
 - Schema(s), standard(s) to enable data exchange between data storage structures. The schema should be based on vocabularies established in the MOD-CO context as well as on existing third-party thesauri
 - Information models as the basis for database architectures
 - Implementations or tools for demonstrating the practicability
 - Establishment and further elaboration of a MOD-CO wiki as a platform for disseminating the above-listed products

- The products of MOD-CO are not in contradiction to already existing schemas by other consortia, but are regarded as useful additions.

→ Meetings and workshops with potential user groups and members of other standardization consortia are envisaged

2: SHORT REPORTS OF THE THREE WORKGROUPS

- The workgroups in Munich (Triebel) and Bayreuth (Rambold) gave reports about the hitherto achieved deliverables. The Bremen group (Glöckner) will start with its contributions to the work packages in August 2014. The achievements have thoroughly been analysed and discussed.
- Workgroup D. Triebel, SNSB, Munich:
 - Design of the MOD-CO wiki as an internet presence of the project as well as for usage as a working platform for project-internal purposes
 - Based on a preliminary version of the wiki, the following changes the participants agreed on the following changes:
 - Replacement of the menu item *Team* by *Contact*
 - *of Implementations* by *Products*
 - *of Schema* by *Schemas*
 - *of Publications* by *Articles and Reports*
 - *of Current Events* by *Events*
 - Restructuring the sidebar
 - Restructuring the *Internal pages* according to the work packages in the DFG proposal
- G. Rambold, University Bayreuth:
 - The group demonstrated a basic structure and classification of the MOD-CO vocabulary to be elaborated (Appendix, Fig. 1). Aside from the aspect of project-level issues (management, responsible) the following four major data domains were distinguished:
 - **Sampling:** collection issues
 - **Sample processing:** preparation, conservation, storage (DNA including extraction and other types of decomposition of the samples)
 - **Sample relating:** treatments, experiments (experimental design)

- **Sample measuring:** object analysing the samples manually or via digitalization, observations (by human or machine), including sequencing of nucleic acid or protein extracts, analysis of the (meta-)metabolome

3. DEMOS OF THE AVAILABLE TOOLS AND DISCUSSION ABOUT THE ASPIRATED MOD-CO AIMS AND TOOLS

- The MOD-CO wiki is designed as a working platform for the three workgroups and provides different features to support this:
 - Contents should be discussed in the discussion tab. It will make it easier for the different responsible persons to consider the views of other project members.
 - The versioning of the products (vocabularies, schemas, models and implementations) should be documented there.
 - Schema versions should be made available as downloadable XML-files.
 - Early versions of the schema should be tested and verified soon by the workgroups using Silva/megx.net (for example) and Diversity Collections/Descriptions.
- Demo of SILVA/megx.net by P. Yilmaz

A dataflow with the DWB components DeltaAccess and DiversityDescriptions v. 3 was proposed by the Bayreuth group to be an adequate joint platform for elaborating the vocabularies and schemas. Stations of an example schema dataflow starting with a descriptor list in the DELTA format – import into DiversityDescriptions v. 2 – export and reimport (in SDD format) into DiversityDescriptions v. 3 and export were demonstrated by using screenshots.

4. TO DOS AND DISTRIBUTIONS OF TASKS

- A (preliminary) joint decision on the acceptance of the four proposed data domains (Appendix, Fig. 1) should soon be achieved.
- Individual descriptors should be documented in a tripartite structure as follows: Element → Attribute → Value/States/Options
 - Given that the three workgroups have agreed on these two issues the elaboration of the namespaces of the data domains has to be initiated: *Sampling* and *Sample relating* will mainly be curated by the Bayreuth group, *Sample processing* and *Sample measuring* by the Bremen group.
 - Preliminary, all four major data domains are curated within one list; subsequently they may be split into four separate lists.
- Additions in section Article & Reports (if and where necessary) should be made continuously by all groups.
- Renaming and restructuring of the internal wiki part and the wiki sidebar menu will be done by the Munich group.

- A next workshop with invited participants is scheduled for end of 2014.
- *GitHub* should be evaluated for usage of a collaborative editing tool aside from *DiversityDescriptions*.
- The purpose of every Wiki page should be explained in a header information. A draft version of these may be elaborated by the Bayreuth group.

APPENDIX

Table 1: Timetable of the DFG supported project MOD-CO

Year	1				2			
Quarter	I	II	III	IV	I	II	III	IV
Project coordination, Wiki administration,	■	■	■	■	■	■	■	■
Internal meetings of the participants	■				■			■
Setup of the MOD-CO Wiki	■							
Set up of the MOD-CO namespace schema	■	■	■	■	■	■		
Set up of MOD-CO assigned standard vocabularies		■	■	■	■	■	■	
Workshops with invited colleagues from the fields of Meta-omics research and Biodiversity data standardization			■			■		
Implementation of the schema into DWB DiversityDescriptions (MOD-CO-DD); experimental import and export of meta-omics example data.						■	■	
Implementation of the schema into SILVA/megx.net; experimental import and export of meta-omics example data.						■	■	
Generation of the MOD-CO final draft schema and standard vocabulary, writing of the final report and submission of the MOD-CO final draft as a standard proposal to GSC/TDWG							■	■

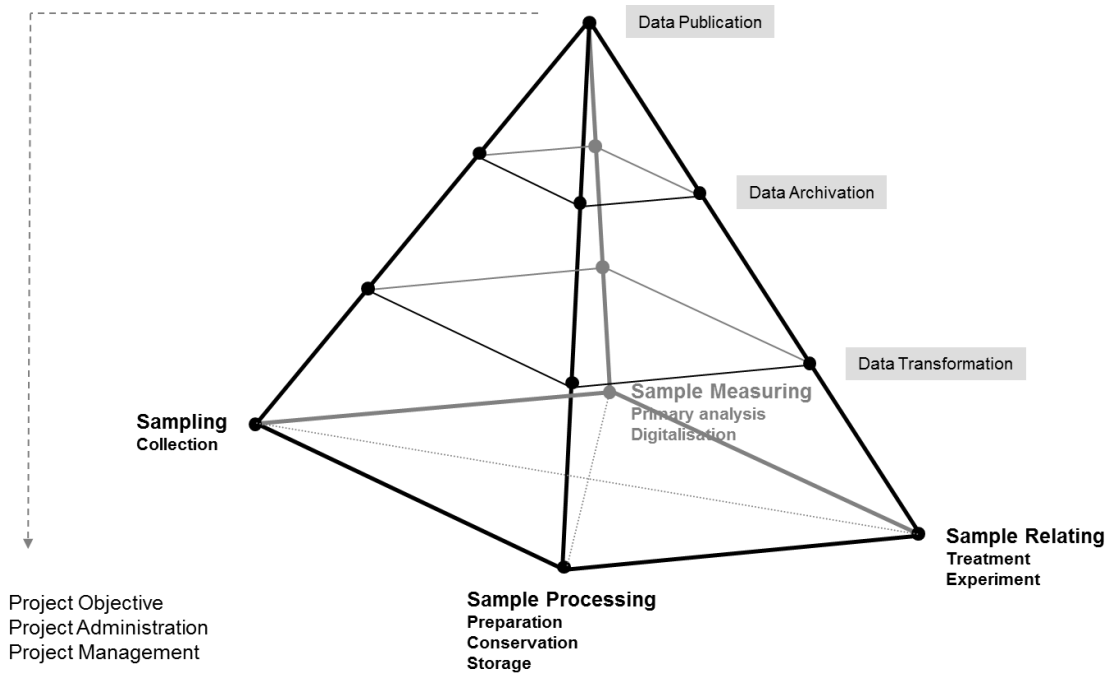


Figure 1: Domains of MOD-CO descriptors