D1.2-DB: Completion of the MP/ML database

_____

Project acronym:	**BONUS MICROPOLL**
Project full title:	**Multilevel assessment of microplastics and associated pollutants in the Baltic Sea**
Deliverable number:	Call2015-122_D1.2-DB
Nature / type of deliverable:	RE - report
Dissemination level :	PU - Public
Delivery month :	14
Work package leader:	**Leibniz Institute of Polymer Research Dresden e.V., Germany**
Authors:	**Dieter Fischer, Natalja Čerkasova**
Last modified:	**2018-09-24**

Objectives of WP1 - Marine MP sampling & processing

Our hypotheses are that highly-populated areas (via estuaries etc.) and Baltic Sea lanes are the most relevant sources for MP, and that beaches and coastal sediments are the major sinks of MP. On the other hand, besides beaches, vertical flux of MP towards the sediments is supposed to be one of the essential sinks in the Baltic Sea. To test this, the aim of this WP is to identify MP distribution:
- vertically, in the water column and associated sedimentation rates.
- spatially, in water, biota, sediment and beaches of relevant estuaries, shorelines and the open seas in the Baltic Sea.

Results

Summary

A Marine Plastic Database (MPDB) was developed. The MPDB is a database for all data collected within the project. It allows general import/export and data exchange with other international and national Marine Litter databases (e.g., German Baltic Marine Litter database) and provides tools for data quality insurance and assessment, spatial visualization, and statistical analysis. The database is developed using SQL as a widely spread and well-documented language with the necessary flexibility of storing data.

Towards the next months and until the end of the MICROPOLL project the database will be further improved and expanded, if all users of the project use the database and give then comments and input.

Database description and aim

MICROPOLL Database (MPDB) is a database for all data collected within the project. It allows general import/export and data exchange with other international and national Marine Litter databases (e.g. German Baltic Marine Litter data base), and provides tools for data quality insurance and assessment, spatial visualization and statistical analysis.

The database is developed using SQL as a widely spread and well-documented language with the necessary flexibility of storing data. Key points of the task:
- DB allows general export and data exchange (csv, ascii, xml);
- DB allows spatial visualization (with extra tools) and statistical analysis;
- An assigned person with sufficient IT skills (Natalja) will handle DB administrative management (thru the project implementation phase). Other users receive a user key (with limitation, to assure data quality and intactness).
- DB should be marinated following a technical maintenance plan which should include:

- Data quality assurance;
- Data backup plan;
- Risk assessment plan;

Database structure

The developed database structure is presented in Figure 1. The main tables are:
• Samples table;
• Analysis table;
• Particles table.

The supporting tables are:
• Colors;
• Shapes;
• Size fraction;
• Polymer category;
• Paint remark;
• Countries;
• Institution;
• Equipment;
• Method_category;
• Projects;
• Sampling_compartment;
• Affiliation.

The many-to-many relationship table is:
• Particles2analysis.

<u>Supporting tables</u>
Information in the supporting tables has to be filled up prior to the main tables. The information to be filled should follow the logic:
1. User fills out his/hers institution in the "Institution" table, if necessary (to be completed only once);
2. User fills out his/hers credentials in the "Contributor" table, if necessary (to be completed only once);
3. Fills out the "Projects" table, if necessary (to be completed only once);
4. User goes thru all the supporting tables and fills out the information needed (to be completed once or multiple times, depending on the situation);
   o Equipment (which equipment was used);
   o Method (which method was used);
The tables "Color", "Shape", "Polymer category", "Paint remark", "Size fraction" are already filled up with the agreed-upon categories.

Main tables

After completing the information input of the supporting tables, the user may proceed to the main tables. The data entry should follow the logic:

- A sample entry should be created in the "Samples" table, which provides the information on the collected sample/samples;
- An analysis method should be created in the "Analysis" table (if not created prior), by which these samples were analyzed;
- The Particles entries should be added to the "Particles" table. The user may specify how many on which category, shape, size, color, etc. particles were found in which sample analyzed by which method.

Relationship tables

Because the basis of the relation database is the one-to-many data integrity and structure, for many-to-many relationships a separate table should be included, which will define this relationship. The table is called "particles2analysis" and defines the IDs of those particles or those particles, who were found by multiple analysis method. In short, this table provides the linkage between multiple particles found by multiple analysis and contains the unique IDs of said entities.

Database access

The users may request the MPDB access via e-mail to: Natalja.cerkasova@gmail.com. Every user is issued with unique credentials for the DB usage.

The DB is currently hosted by the IOW server at 192.124.245.26 (micropoll.io-warnemuende.de).

The instructional videos for MPDB usage are available at google drive storage (link): https://drive.google.com/drive/folders/1hdwP6AmJ1AAYQ-JbPyR0Sf7XGZGH_I5Z?usp=sharing

Future work and fixes

Future work includes:

- The assessment of circular references within the database and possible separation of the tables;
- The assessment of fields and their constraints/requirements.

Figure 1. MPDB EER diagram

analysis
- IDAnalysis INT(10)
- Spectrum LONGBLOB
- Library_entry VARCHAR(50)
- Hit_quality FLOAT
- Result TEXT
- Comment TEXT
- Indexes

particles2analysis
- IDParticles INT(10)
- IDAnalysis INT(10)
- Indexes

affiliation
- IDAffiliation INT(10)
- affiliation VARCHAR(50)
- Indexes

colors
- IDColor INT(10)
- Color VARCHAR(50)
- Indexes

shape
- IDShape INT(10)
- Shape VARCHAR(50)
- Indexes

size_fraction
- IDSize INT(10)
- Size_category VARCHAR(50)
- Size_range_from FLOAT
- Size_range_to FLOAT
- Indexes

polymer_category
- IDPolymer INT(10)
- Polymer_category VARCHAR(50)
- Indexes

paint_remark
- IDPaint INT(10)
- Paint_remark VARCHAR(500)
- Indexes

particles
- IDParticles INT(10)
- Sample VARCHAR(50)
- Method INT(10)
- Arrival_date DATE
- Analysis_date DATE
- Amount FLOAT
- Analyst INT(10)
- Size_fraction VARCHAR(50)
- Colour VARCHAR(50)
- Shape VARCHAR(50)
- Size_1 FLOAT
- Size_2 FLOAT
- Image LONGBLOB
- Categorised_result VARCHAR(50)
- Indication_paint VARCHAR(50)
- Indexes

contributor
- IDContributor INT(10)
- Fisrt_Name VARCHAR(50)
- Last_Name VARCHAR(50)
- Initials CHAR(50)
- Institution INT(10)
- Affiliation VARCHAR(50)
- Contact_phone VARCHAR(50)
- Contact_email VARCHAR(50)
- Institution_other VARCHAR(50)
- Indexes

samples
- IDSample INT(10)
- Sample_name VARCHAR(50)
- Project VARCHAR(50)
- Contributor INT(10)
- Method INT(10)
- Site_name VARCHAR(500)
- Country VARCHAR(50)
- Compartment VARCHAR(50)
- Date DATE
- Time TIME
- Sampling_duration TIME
- Site_comments TEXT
- GPS_LON FLOAT
- GPS_LAT FLOAT
- GPS_LON_end FLOAT
- GPS_LAT_end FLOAT
- GPX_track MULTIPOINT
- Sampling_depth FLOAT
- Sampling_distance FLOAT
- Sampling_area FLOAT
- Sampling_weight FLOAT
- Sampling_species TEXT
- Purification_date DATE
- Fraction_analysed FLOAT
- Analysis_completed DATE
- Indexes

sampling_compartment
- IDCompartment INT(10)
- Compartment VARCHAR(50)
- Indexes

projects
- IDProject INT(10)
- Acronym VARCHAR(50)
- FullName TEXT
- LeadingInstitute INT(10)
- Indexes

countries
- IDCountry INT(10)
- CountryShort VARCHAR(50)
- Country VARCHAR(50)
- Indexes

methods
- IDMethod INT(10)
- Method_name VARCHAR(150)
- Category VARCHAR(50)
- Equipment INT(10)
- Specification TEXT
- Scheme_or_procedure TEXT
- Images LONGBLOB
- Protocols LONGBLOB
- Equipment_owner INT(10)
- Indexes

institution
- IDInstitute INT(10)
- Institute_Name TEXT
- Institute_Acronym VARCHAR(50)
- InstituteCountry VARCHAR(50)
- Institute_Address TEXT
- Institute_WebPage TEXT
- Institute_Phone TEXT
- Institute_Email TEXT
- Indexes

equipment
- IDEquipment INT(10)
- Eq_Name VARCHAR(50)
- Eq_Specification VARCHAR(50)
- Indexes

method_category
- IDMethodCategory INT(10)
- Category VARCHAR(50)
- Indexes