

Article

A Public Database for Microplastics in the Environment

Natalja Čerkasova ^{1,2,*} , Kristina Enders ^{3,4} , Robin Lenz ^{3,4} , Sonja Oberbeckmann ⁴ , Josef Brandt ⁵ , Dieter Fischer ³ , Franziska Fischer ³ , Matthias Labrenz ⁴  and Gerald Schernewski ^{1,4,*} 

¹ Marine Research Institute, Klaipeda University, Universiteto ave. 17, LT-92294 Klaipėda, Lithuania

² Texas A&M AgriLife Research, 720 E Blackland Rd, Temple, TX 76502, USA

³ Leibniz Institute of Polymer Research Dresden; Hohe Straße 6, D-01069 Dresden, Germany

⁴ Leibniz-Institute for Baltic Sea Research, Seestrasse 15, D-18119 Rostock, Germany

⁵ Department of Marine Sciences, University of Gothenburg, Box 461, SE-405 30 Göteborg, Sweden

* Correspondence: natalja.cerkasova@ku.lt (N.Č.); gerald.schernewski@io-warnemuende.de (G.S.)

Abstract: During recent years plastics became one of the focuses of EU policy. A harmonisation and comparability of microplastics monitoring results across Europe is needed. The complexity of microplastic data makes it necessary to develop a specific, tailor-made database rather than adapting and modifying one of the existing databases. To meet this demand, we present a publicly accessible, flexible, and extendable structured relational database for particle-based microplastic data. The developed relational database is adaptive and meets the specific demands of microplastics, e.g., a large variety of sampling, processing and analytical methods, many types of plastics, and a very wide size spectrum ranging from micrometres to millimetres. In this paper we discuss the development of the database, data entry specifics, sample analysis methods, microplastics data manipulation and quality assurance, and database integration and accessibility.

Keywords: microplastic data; database; marine; particles; EU policies; Marine Strategy Framework Directive



Citation: Čerkasova, N.; Enders, K.; Lenz, R.; Oberbeckmann, S.; Brandt, J.; Fischer, D.; Fischer, F.; Labrenz, M.; Schernewski, G. A Public Database for Microplastics in the Environment. *Microplastics* **2023**, *2*, 132–146. <https://doi.org/10.3390/microplastics2010010>

Academic Editors: Nicolas Kalogerakis and Tony Robert Walker

Received: 30 December 2022

Revised: 15 January 2023

Accepted: 7 February 2023

Published: 15 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Environmental policy in European countries is driven by ambitious European Union (EU) legislation, which has affects beyond the EU member states. Some major EU policy fields are the marine and coastal environment, nature and biodiversity, and water. The water policy aims at protecting the European Union's water resources and ecosystems and ensuring access to clean drinking and bathing water. This includes specific policies, e.g., on marine, surface, bathing and drinking water, floods, nitrate pollution, groundwater, urban wastewater, water reuse, water scarcity, and droughts. Most environmental policies are implemented though legally binding directives. These directives are accompanied by implementation guidelines that regulate, for example, monitoring, methodological standards, quality thresholds, and environmental quality assessment. A consequence of EU environmental policy implementation is the collection of enormous amounts of heterogeneous environmental data.

The need to deal with this data initiated the development of the EU Shared Environmental Information System (SEIS). It was established to improve the collection, exchange, and use of environmental data and information across Europe [1]. The SEIS serves as an umbrella for specific data platforms such as the Marine Information System for Europe (WISE-Marine) and regional sea platforms, e.g., for the North Sea (OSPAR), or the Baltic Sea (HELCOM). WISE-Marine and the regional platforms directly support European environmental policies' implementation. These international approaches are complemented by national environmental data platforms, for example, in Germany the Marine Data Infrastructure Germany (MDI:DE).

The general marine data management in Europe consists of three major elements: SeaDataNet, EMODnet, and Copernicus CMEMS. SeaDataNet is a distributed marine data

infrastructure for diverse in situ data sets. EMODnet is using available data for preparing multi-resolution maps of all European seas, and Copernicus is the earth observation program and offers information services mainly based on satellite data.

It is obvious that the large amount of heterogeneous data, multiple platforms that depend on data-exchange, and a wide range of potential data uses require rules that ensure data quality and accessibility. These rules are provided by the EU INSPIRE Directive. An important aspect is that INSPIRE contains guidelines on metadata documentation based on EN ISO 19115 and EN ISO 19119. Metadata provide the required background information about data. Metadata descriptions are a prerequisite so that data, physically stored in a database, can be found and used by others.

During recent years, plastics became one focus of EU policy. In 2018, the EU adopted the plastics strategy (EC COM/2018/028, 2018) with the aim to protect the environment and reduce marine litter. It is part of the circular economy action plan, revised in 2020, and builds on existing measures to reduce plastic waste. In the marine environment this is complemented by the EU Marine Strategy Framework Directive (MSFD 2008/56/EC). The MSFD has the aim to protect and achieve a good environmental status (GES) in marine ecosystems. The GES is assessed on the basis of eleven descriptors and marine litter is one of them, which aims to ensure that properties and quantities of marine litter do not cause harm to the coastal and marine environment. As a consequence, the state-of-pollution of the marine environment has to be assessed, environmental targets and associated indicators have to be developed, major emission sources have to be identified and quantified, and effective measures leading to reductions in marine litter pollution have to be tested. One focus is on trends in the amount, distribution, and composition of microplastics. This has caused a wide range of scientific and monitoring activities during the last decade and the need to establish joint databases that follow FAIR (Findable, Accessible, Interoperable, and Reusable) data principles [2].

A first European litter database, linked to EMODnet, exists [3]. It focuses on macro-litter, the size fraction above 25 mm, found at beaches using the 100 m monitoring method [4]. A recent review shows that the database suffers from many data weaknesses, for example, spatiotemporal heterogeneity, inconsistencies in the different litter identification lists, and variability in the survey protocols and data sources. The conclusion is that a harmonisation and comparability of results across Europe are still needed. For microplastics, a comparable database does not exist yet, but there are several lessons learned from the macro-litter beach database. A public microplastic database should be available at an early stage of monitoring. It needs to be flexible and adaptive and must meet the specific demands of microplastics, e.g., a large variety of sampling, processing and analytical methods, many types of plastics or a very wide size spectrum ranging from micrometres to five millimetres. The complexity of microplastic data makes it necessary to develop a specific, tailor-made database rather than adapting and modifying one of the existing marine databases. This means that the new database needs to be visible for and integrated into existing database structures.

To meet this demand, we present a publicly accessible, flexible, and extendable structured database (DB) for particle-based microplastic data, that meets the FAIR data principles, takes into account the INSPIRE metadata requirements, and is linked to and harvested by other databases.

2. Materials and Methods

2.1. Basic Requirements and Background

The database was developed to meet the data storage and reporting requirements of two international projects with focus on coastal and marine microplastics in the Baltic Sea region, namely Bonus MicroPoll (<https://www.io-warnemuende.de/project/192/micropoll.html>, accessed on 12 December 2022) and MicroCatch (<https://www.io-warnemuende.de/microcatch-home.html>, accessed on 12 December 2022). Database requirements were that relevant sample parameters had to be efficiently (i.e., no data loss, consistency, simplicity)

communicated between the entities. It had to be possible to reconstruct each individual microplastic particle's history from sampling to final analysis. This was the key for ascertaining data quality in terms of good scientific practice. It had to be possible for sample recruiters to share sample locations, sample type, sampling method, and time. Contact details of the responsible persons of each sub-task had to be taken into account. Important information concerning the sample purification, such as treated sample amount, blanks (contamination control), purification methods, and date had to be represented. Among relevant parameters of microplastic analysis were, e.g., polymer type, size, particle shape, colour, analysis method, the fraction analysed (if a purified sample was split), and blanks.

All the above information and additional required details were implemented into the Marine Plastic Database (MPDB) design and optimised over several years with respect to scientific accuracy, relevance, clarity, and consistency of terms and information within and across database tables. In particular, the two latter requirements were driven by the goal to make the MPDB publicly available. Therefore, we aimed at reducing the information to a minimum that is mandatory for marine plastic-monitoring in the environment. At the same time, it was important to keep flexibility with respect to the large variability of different sample requirements (e.g., concerning purification or analysis methods). The database had to take into consideration that microplastic monitoring involves different actors across various stages with different tasks, such as sample collection, sample purification, and spectroscopic analysis.

2.2. Types of Data

Quantitative microplastic (MP) data is typically either mass- or particle-based, both methods having their own advantages and limitations. Eventually, the targeted polymer types and size ranges, and the foremost research question determines the analytical method of choice.

Mass-based data, as for example from pyrolysis gas chromatography mass spectrometry (Py-GC-MS), usually necessitates less sample purification and hence accelerates data generation. Losses caused during pre-treatment, which particularly affect the small size classes, become obsolete and make this analysis especially advantageous for target particles in the sub-micrometre range. Limitations arise on the detection of multiple polymer types, as interferences can occur [5] that involve increasing resources and efforts when aiming at a large coverage of materials addressed by the umbrella term "microplastics". Microplastic particles' polymer types that appear in low amounts are prone to fall below the limit of detection if they belong to a rarely occurring type, or if the particles are so small that they, although being numerous, do not provide enough mass [6,7]. In any case, the main disadvantage is that only the mass of the MP polymer type is obtained and there is no information on particle sizes and their distribution.

Any research question that requires a higher resolution of particle size, examples are the determination of transport patterns, normalisation of biofilm coverage, or the encounter rates of potentially affected organisms necessary for a respective risk assessment, will demand or consider particle-based data acquisition. Particle-based analysis is usually performed using μ -Raman- or Fourier-Transform-Infrared-Microspectroscopy (FTIR) with the ability to detect very low numbers of polymer particles. The methods are either done in imaging or particle measurement mode. Imaging entails scanning the entire filter area and acquiring spectra at every pixel. After spectra evaluation, adjacent pixels with the same chemical classification are grouped as particles [8,9]. In particle measurement, at first an optical image of the filter is acquired, from which the location of the individual particles is determined, together with information about size, shape, and colour or texture. Only then are the spectra acquired at the locations where particles were found [10,11].

2.3. Morphological Properties of Microplastic Particles

The minimum size of particles to be detected depends on the spectroscopy technique used (FTIR: approx. 10–15 μm , Raman: approx. 1 μm). Which technique is used also de-

depends on the sensitivity of the analysis towards certain substances [12]. Spectra quality, and thus, method reliability can be impacted by fouling of particles or fluorescent components (in case of Raman).

Using the particle-based approach, concentrations are typically reported as MP counts per unit environmental matrix (e.g., kg^{-1} , or m^{-3}). In addition, other particle characteristics, such as shape properties or colour can be determined on a per-particle basis.

The particle count concentrations themselves are, however, of limited usefulness when being reported without substantial context. As an example, estimates of how many trillion “pieces” of MP are polluting the ocean, are readily incorporated and circulated by various media [13,14]. It is questionable if actual informative content about the pollution’s extent can be conveyed in this manner, or rather, if a broader public attraction is intended simply by the magnitude of an unimaginable high number. Representations of particle sizes (e.g., as individual particle properties or continuous distributions [15]) and study design details (e.g., applied mesh sizes) are examples of such augmenting data that should preferably be communicated alongside MP concentrations. The same is true for comparative data on natural particles where applicable (e.g., as proxy for the hydrodynamic regime in sediment MP studies [16]). An appropriate database structure for particle-based MP pollution data should therefore provide the possibility to map all relevant morphological data which is then available for any future assessment. From simple to sophisticated, such morphological properties (“morphologicals” hereafter) may include size class [16,17], longest size dimension, [18] or multiple size dimensions, area [19,20], circumference, or shape related aspects, such as sphericity, convexity, eccentricity, right up to fully 2D or 3D descriptions of the irregular particles (e.g., as point meshes). Referenced studies in the above list give examples from the scientific literature that utilised the respective morphological features to describe their identified particles. The more sophisticated morphological features may have higher demands for the analytical process, but they give the resulting data set a higher diagnostic quality for describing the interactions the identified particles may invoke in the environment. Although not trivial, particle mass estimates may be inferred if the reported particle morphological data are of sufficient detail, i.e., at least a proper shape classification and maximum extents in 3 dimensions. This allows to link the realms of particle- and mass-based analyses, which would be highly beneficial to the field of MP research. In contrast, inferring morphological data from destructive mass-based analytics is unfeasible.

Fortunately, for the choice of appropriate morphological features, it is not necessary to start *ab initio*, as much knowledge can be transferred from other particle-based research fields, such as from sedimentology, with a great repertoire on various size characteristics of sediment grains [16,20,21].

2.4. Software and Tools

Prerequisite for a database is that it contains a structured set of data, meaning that data should be organised using formal design and modelling techniques. Structured data, which usually reside in relational databases, provide benefits, including but not limited to data integrity and independence from applications or programs, increased consistency, decreased updating errors, improved data access to users, and data security. This ensures that the research data meet the FAIR principles—findable, accessible, interoperable, and reusable. It is worth noting that the mentioned advantages entail additional expenses such as the time and intellectual requirements to design a complex database, possible software and hardware costs, and costs for training the personnel to use and administrate, as well as maintain the database.

An SQL (structured query language) database can be effectively set up to represent multiple levels of abstractions. It is a collection of highly structured tables, wherein each row reflects a data entity, and every column defines a specific information field. Relational databases are built using the structured query language to create, store, update, and retrieve data.

The MySQL™ software delivers a very fast, multithreaded, multi-user, and robust SQL database server [22], hence was considered to be a perfect tool for the development of the MP database. The MySQL software is Dual Licensed; in this study the MySQL as an Open-Source product under the terms of the GNU General Public License (<https://www.fsf.org/licenses/>, accessed on 10 December 2022) was used. Depending on user preferences and requirements, other tools were in use as well, such as the HeidiSQL, [23] MySQL Workbench (<https://www.mysql.com/products/workbench/>, accessed on 10 December 2022), and certain Python libraries (mysql.connector [24]) for automated access to the DB. Any relational database management system (RDBMS) can be used to access the database, as long as it uses SQL as its standard database language. The database server is hosted by the Leibniz Institute for Baltic Sea Research Warnemünde-IOW, which allows usage by registered users. As in any database management system (DBMS), users may perform tasks which correspond to their rights, granted by the administrator: view, edit, delete, etc. The possible database usage possibilities are explained in MySQL documentation (available at <https://dev.mysql.com/doc/>, accessed on 10 December 2022), and many tutorials and videos present online [25].

2.5. Database Development

The database development process was conducted following a classical waterfall model [26], with iterative process revision (Figure 1). Firstly, specific needs for data were identified, followed by determination of data types. The first step was carried out mainly by the environmental scientists, who oversaw collecting or analysing samples; judging from their experience, these users could immediately identify what data should be collected, whether it is possible to collect all the data (obligatory fields in the database), which data is optional or incidental (optional fields), and what could be further used in later research (quantification of entries). This step was followed by a conceptual model development, which established connections between possible tables and entities. The conceptualisation was completed with tight communication between the team members, who oversaw sampling and analysis, and those responsible for the MPDB technical implementation (database developer). Where one could identify the relationships between data, the other could comment on the possibility of implementing such a relationship and to assure database structure logic.

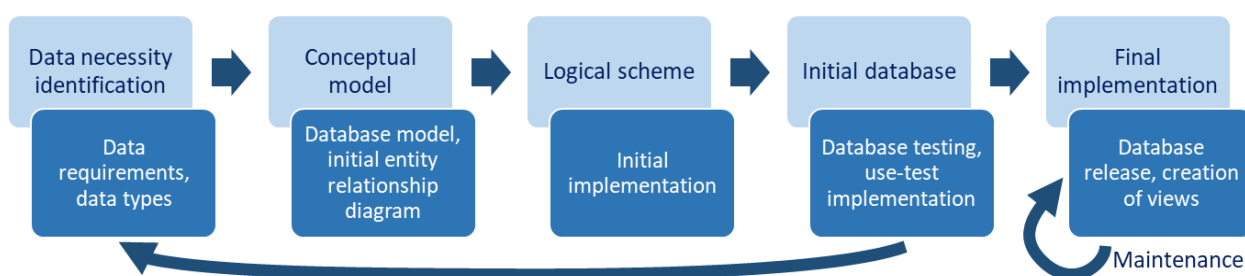


Figure 1. Marine Plastic Database development process stages.

Further steps (logical schema, initial and final implementation) were completed by the developer, based on the possible use-case scenarios, provided by the MPDB users. Currently the database is in its finalised stage, with maintenance and the support to create custom database views on demand. Database views are pre-established query commands, which are kept in the database dictionary to generate a certain preview of structured data, possibly combining entries from multiple tables or other views, performing certain calculations, or simplifying multiple tables to a single table, which can then be used for further data analysis. The use of such views greatly facilitates data extraction for specific repetitive evaluation tasks.

The MPDB development stages 1 to 4 (see Figure 1) were revisited several times, as some requirements and improvement suggestions were received after the initial test-

ing of the database, which in turn assured a reliable product that suits the users and stakeholder needs.

2.6. Data Entry

The most efficient method of data entry depends on the type of data, its amount, and the frequency with which it accrues. Some data entries are best entered manually, e.g., to populate the ‘contributors’ table (a contributor is a person who is working with the samples or analysis), or some categories, such as the ‘colour’ or the ‘shape’ (we use the term “auxiliary” tables, when describing such entity properties). Such tables are usually referred to as pre-populated categories. For bulk data entry, however, a manual approach would be counter-productive and undermine the development of the relational database in the first place. Instead, the database ought to be populated wherever possible with bulk entries using queries, spreadsheets (comma-separated values (CSV) sheets), any structured text document, as well as using fully- or semi-automated scripts.

Data entry for the ‘samples’ table were most useful in the form of SQL ‘insert’ queries which pulled a prepared CSV file into the table. The creation of data describing samples typically involves highly manual tasks, such as insertion of information from handwritten field work logbooks. Most researchers will prefer a simple spreadsheet application for this work. In such a case, a simple export of the DB table into a CSV file format may be utilised and further used as a template where the new data can be appended. This ensures that the later CSV upload to the DB can be completed by a standard SQL script without errors or data incompatibilities in a semi-automated way.

The tables containing actual analytical result data (such as ‘particles’ or ‘analysis’) receive the highest number of entries (as there may be hundreds or thousands of MP particles per sample) and are therefore impractical to be populated by manual data entry or individual CSV files. For this purpose, a Python software package was developed which performs an automated particle measurement and data evaluation and entry [27]. Using the `mysql.connector` Python library [24] allowed streamlining the necessary data compilation, conversion, and SQL insertion to reduce manual effort to a minimum. The custom-made graphical user interface (GUI) allowed reviewing summarised and categorised particle information and uploading the results of hundreds of particles in less than a minute.

3. Results

3.1. Database Structure

The final structure of the MPDB is presented in an enhanced entity-relationship (EER) database diagram which represents requirements and complexities of the developed database (Figure 2). There are 14 auxiliary tables, 3 main, and 2 junction tables. The data in the auxiliary tables have to be entered only once or updated very rarely, i.e., if new contributors (a person who is working with the samples or analysis) are engaged in the procedure, then their information should be entered once. Every contribution (sample or analysis) will be automatically assigned to that person with a unique ID. Hence, the information is stored once and is referenced via a foreign key attribute.

The three main tables of the database are: samples, analysis, and particles (see also Figure 3). Any sample may be analysed via various methods and yield many particles. For instance, a certain sample was treated with several different purification methods, all of which had to be registered in the database. Similarly, a certain particle could be measured with both methods, FTIR and Raman, yielding individual results that should all be included in the MPDB. Hence, the standard relational database rule of one-to-one or one-to-many relationship is not sufficient; a many-to-many connection had to be established. This is achieved via the two associative or junction tables: ‘sample2methods’ and ‘particles2analysis’. The connections are stored in these junction tables via the copies of the primary keys of joining tables (Figure 2).

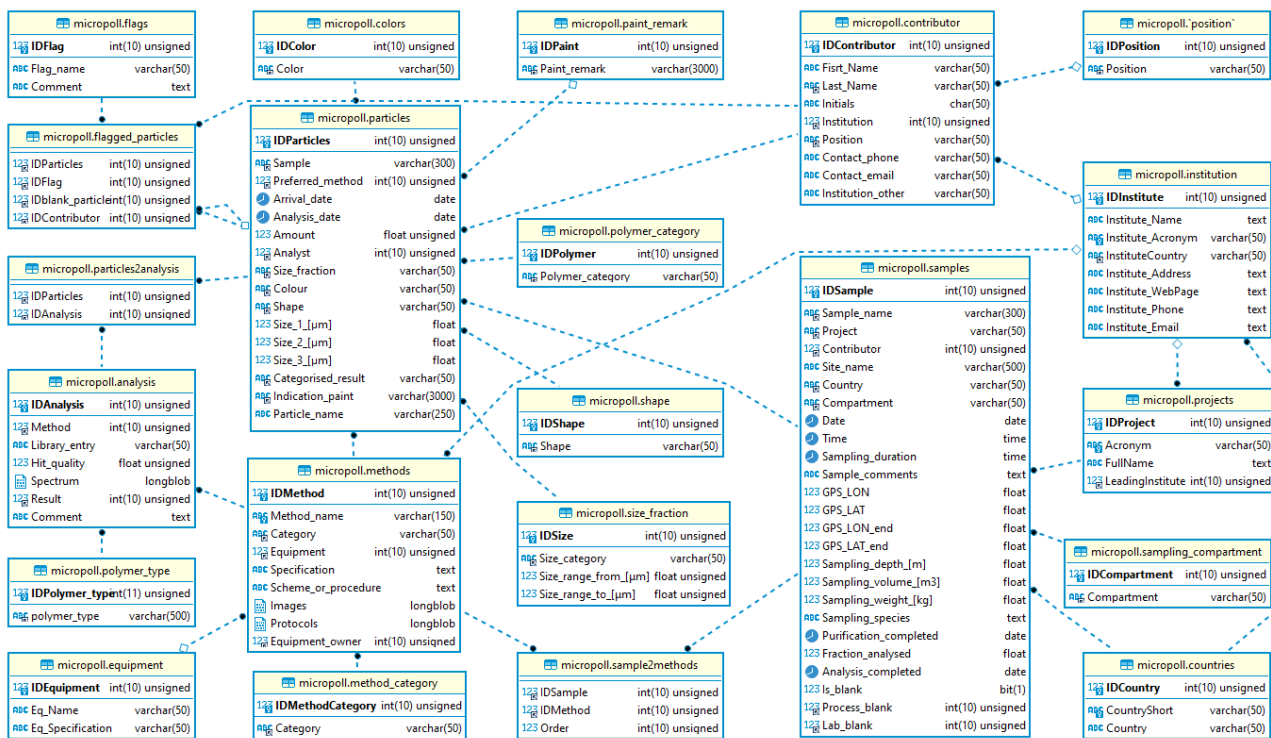


Figure 2. Enhanced entity-relationship diagram of the final version of the Marine Plastic Database.

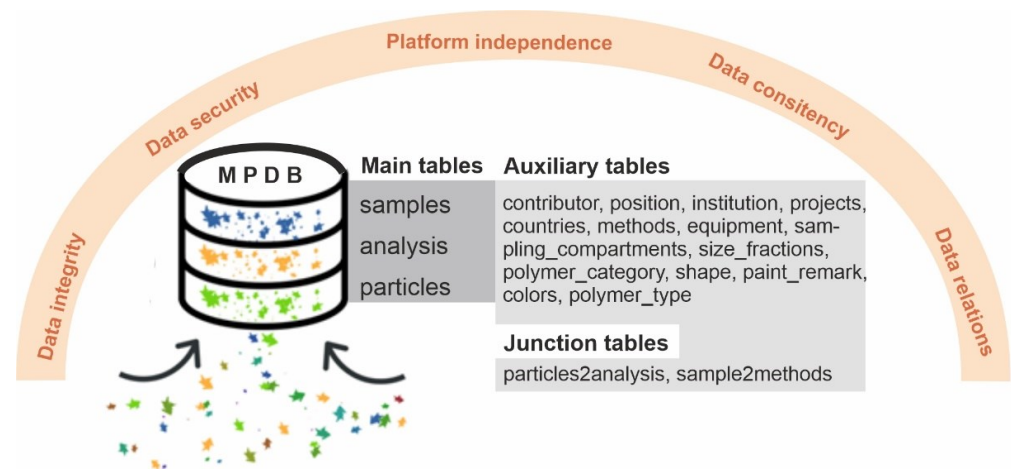


Figure 3. The Marine Plastic Database with main and auxiliary tables as well as requirements and advantages.

The database users usually work with the main tables and the junction tables. The unique ID keys are assigned automatically. The ‘samples’ table receives entries for actual environmental samples as well as blank samples for contamination control. The field ‘Is_blank’ requires a Boolean-type entry to unambiguously mark the type of each sample. We use two different kinds of blanks: one covering the particle analysis steps (called lab blanks) and others covering the sample purification (called process blanks).

Lab blanks represent potential contamination introduced during analysis preparation and spectroscopic measurement (FTIR or Raman microspectroscopy) of a sample, which is a critical step for introducing contaminations. There is one corresponding lab blank for each environmental sample. Process blanks cover the sample type-specific treatment pipeline [28], where a group of process blanks typically represent potential contamination

from the purification procedures of all environmental samples treated alike during a certain time period.

Based on the 'Is_blank' flag, and the entries for 'Process_blank' and 'Lab_blank' an extraction of either type of samples (environmental sample, process blank, lab blank) can be achieved easily by querying the wanted sequence. Particles found in a blank may be used to adjust the particle numbers of their corresponding environmental samples. A recommendation of how to implement a blank correction in the particle domain is given in Section 3.4. A recent review on blank procedures [29] evaluates several further options.

3.2. Data Extraction and Example Queries

Database data retrieval for further analysis is usually done via the data export of views. Database users may use the existing views or ask the administrator to create views based on their description.

3.2.1. Data Filtering Scenario

Some data may need only a simple filtering from one specific table. A simple data filter example may be as follows:

"every sample, from the MicroCatch project from compartment "water" from location between 12.17–12.2 and 53.95–55.0 (lat, long) in the year 2018".

The SQL query snipped for the example is:

```
SELECT * FROM 'samples'
WHERE
samples.Project = 'MicroCatch' AND
samples.Compartment = 'water' AND
year(samples.Date) = 2018 AND
samples.GPS_LON >= 12.17 AND samples.GPS_LON <= 12.2 AND
samples.GPS_LAT >= 53.95 AND samples.GPS_LAT <= 55.0;
```

3.2.2. Data Retrieval from Two Tables Scenario

When extraction of data from multiple tables is needed, the best solution is to create a new joint table, which is called a View, where all the related data entries are linked. After this, the new table can be filtered as in the data filtering scenario. An example for this scenario can be a modified query, as follows:

"every particle, from location between 12.17–12.2 and 53.95–55.0 (Lat, long) in 2018, with a size fraction of 11–100 µm".

Creation of view:

```
CREATE VIEW 'particles_v_samples' AS
SELECT p.'Sample',s.'Date', s.'GPS_LON', s.'GPS_LAT',p.'Size_fraction'
FROM 'particles' AS p
INNER JOIN 'samples' AS s
ON p.'Sample' = s.'Sample_name';
```

Filtering the view:

```
SELECT * FROM micropoll.particles_v_samples
WHERE
year(particles_v_samples.year) = 2018 AND
particles_v_samples.GPS_LON >= 12.17 AND particles_v_samples.GPS_LON <= 12.2 AND
particles_v_samples.GPS_LAT >= 53.95 AND particles_v_samples.GPS_LAT <= 59.0 AND
particles_v_samples.Size_fraction = '11-100';
```

3.2.3. Data Extraction to a Local File Scenario

The data may be extracted to a spreadsheet or other formats for further analysis. Extraction requires either the usage of DBMS tools (like the export tools available from MySQL workbench, or HeidiSQL), or the direct application of SQL commands:

```
INTO OUTFILE 'D:\full_path\samples_table.csv'
```


FIELDS TERMINATED BY ','
ENCLOSED BY ''''
LINES TERMINATED BY '\n';

The provided examples present only a fraction of the user possibilities and may be modified, i.e., to include only unique values of entries, or additional fields or filters. As the stored data has a spatial locator (X and Y coordinates), one can, for example, utilise the functionality of various GIS tools and plugins to plot and represent the data on maps.

3.3. Data Quality Assurance—Dealing with Different Particles

Microplastics—being particulate pollutants that are typically investigated individually for properties such as material, colour, shape, or size—generate data on a per particle basis (at least for quantitative particle-based studies). For statistical analysis, interpretation and dissemination to the general public and stakeholders, data are typically transferred to aggregated information, e.g., counts of particles or averages of sizes, which are reported per sample, or time, or region. This section aims to elucidate the importance of a transparent handling and dissemination of the raw particle data as a means for better data compatibility and re-usability among MP studies. The understanding of particle-based vs. aggregated data is crucial to be able to perform data operations, such as correcting for self-contamination (blank correction) or limit of detection (to be implemented for removal of singular erroneous or misfit results) as explained in the examples below.

Conceptually, we distinguish between data in the “sample domain” (sampling meta-data such as sample quantity, location, dates) as opposed to the “particle domain” (individual properties of each particle). Particle domain data (PDD) can be used to generate sample domain data (SDD) by aggregation. Sums, means, or other descriptive statistics of PDD from a particular sample or sample group may be aggregated to yield the respective SDD. SDD, in turn, cannot be decomposed into the underlying PDD because the information on individual particles is lost in aggregation.

Figure 4 illustrates the distinct representations of data that are handled by the MPDB: in PDD the base of a single entry is an individual particle and it holds particle property data; whereas in SDD the base is a sample, which has sample metadata and aggregated particle data. The phenotyping procedure (PhT) is used to conduct operations on PDD and translate them into SDD.

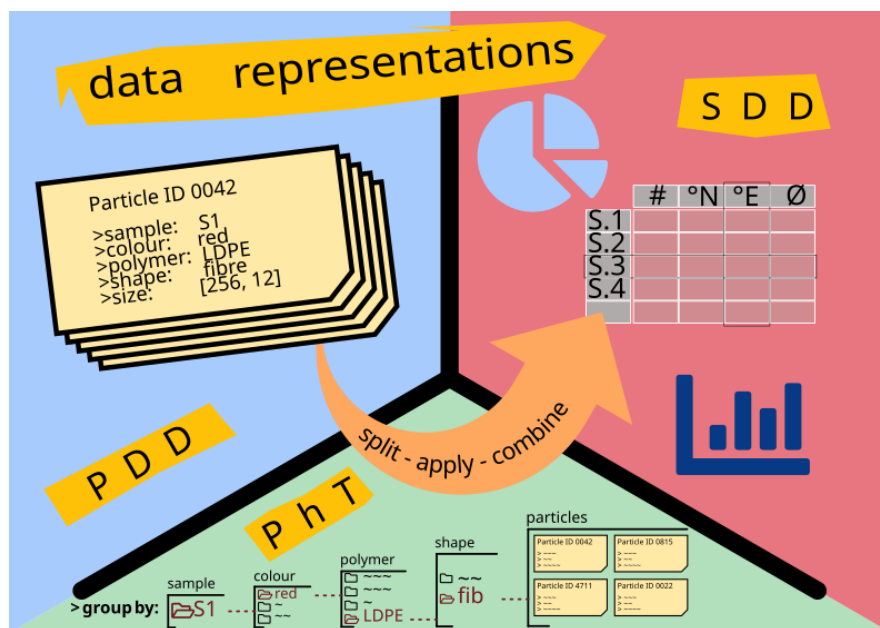


Figure 4. The distinct representations of data that are handled by the MPDB and the algorithms we use are illustrated in this info graphic.

To elaborate the implications of the distinction of SDD and PDD, let us consider the concentration of MP of a sample. In SDD this would be a number of MP per reference quantity of the environmental matrix (e.g., 100 m^{-3} or 23 kg^{-1}). Here, each MP within that number is considered equal, as there are no individual particles regarded by SDD, only their sum. If the underlying raw data (PDD) would only consist of identical particles, a value correction would seem intuitively feasible by a simple subtraction of the required amount. However, MP samples usually do not consist of a number of homogeneous MP particles, but for example, of a red PET fibre and a transparent PE fragment, which can only provide information on possible sample contamination of their kind and not of MP in general. The blank correction must, therefore, be done in PDD, where each particle is *not* equal. In PDD, a subtraction in a mathematical sense is, however, not possible, as each datum represents an individual, not an aggregation. Instead, an elimination of individuals can be an approach to facilitate the correction.

In case of negligible contamination, this theoretical excursion into particle identity should be of little relevance for reported MP data. However, experience shows that (1) contamination is an issue of concern, especially for studies working with small MP, and (2) not seldomly, MP data sets are on the brink of statistical exploitability due to low numbers of the costly to identify specimens, therefore data handling and sanitation procedures are to be chosen carefully.

The here described implications of working with particle data versus aggregated data does not only apply to MP research. Approaches of particle-based data conversion are, for example, discussed in polydisperse systems modelling [28].

3.4. Data Sanitation-Phenotyping and Blank Subtraction

As stated above, the omnipresence of MP easily leads to self-contamination of MP samples, which necessitates blank sample corrections. As the raw data of particle-based MP studies is represented in PDD, a number of individual particles needs to be selected for being regarded as introduced by the sampling and lab procedures, and therefore rejected from final results. The decision of which particle needs to be eliminated must be based on a set of rules. Naturally, there will be no perfect correspondent among the individuals, so the rules need to be designed to find the best matching pair of a blank and an environmental MP particle. Although we know what particles were present in the environmental and respective blank samples, the result will differ depending on the selected elimination rules and methods. To facilitate a matching procedure between 'non-equal' particles, we propose a phenotyping procedure that aggregates all relevant categorical particle properties and uses size as single continuous measure for discrimination. A phenotype is a group of particles that share the same values for 'sample', 'colour', 'shape', and 'polymer type'. It is recommended to unify the categories of colour and shape into few, unambiguous clusters, i.e., red, green, blue, other, or fibre, non-fibre. More detailed specifications of phenotypes come along with a higher uncertainty (e.g., due to differences in illumination or particle shape recognition), especially when working with small MP. Particle size may be converted into a single measure, if not already so, i.e., in some cases we receive a major and a minor length for each particle (first and second dimension), where we take the geometric mean of the two as it is more area-preserving than the arithmetic mean. Phenotyping provides groups of particles which are both SDD (number of particles of phenotype X per sample Y) as well as PDD (the actual particle data in each phenotype group including the single measure size attribute).

It now becomes possible to conduct a blank correction process in PDD by identifying the corresponding blank for the sample in question and consecutively matching the phenotypes present in both. Particle elimination may then be conducted by matching and removing the nearest sized sample particle for each blank particle one by one. In a case where multiple (process-) blanks were taken to cover the same work process they would be, together as a group, valid for all samples prepared by that process. Consequently, for a certain phenotype, the average number of particles per blank is equal to its expected value

of contamination from the work processes covered by the blanks. However, averaging is a SDD operation and cannot be easily performed in PDD. Therefore, we recommend the following procedure/algorithm:

- After phenotyping (as described above) the lists of particles are—for each phenotype separately—concatenated from all blanks;
- The combined list is sorted by size;
- It then needs to be shortlisted to only yield the number of particles required by the SDD average (hence, total number of particles of the respective phenotype in all blanks divided by the number of blanks and rounded to the next integer);
- To achieve minimal size distribution deformation, the particles would best be selected beginning from the median of the sorted combined list and selecting every n -th element towards smaller and larger sizes alternately until k particles are selected, with n being the number of blanks and k the respective phenotype's SDD average;
- The shortlisted group of particles may be used further for the actual blank subtraction procedure on the sample particles in the same way as if there was only one relevant blank available.

Rounding, stereotyping particles into phenotypes, and the use of a unified size measure are assimilations that unavoidably introduce distortions, leading to an apparent, systematic error in the results. After all, every particle is *not* equal.

3.5. Data Manipulation and Quality Assurance

Depending on the analysis setup, it may be adequate to eliminate rarely occurring outliers from the data set, such as unlikely polymer types, which may otherwise elicit wrong-headed conclusions. For example, it might be decided that for a particle phenotype that occurs only once in the entire data set or in a sample, there is not enough analytical support that this is a valid particle, and it should thus be excluded as an outlier. This concept is also known as limit of detection and limit of quantification and applied in many other analytical research branches [30–32] in order to increase the soundness of a particular dataset. In MP analysis, such rare occurrences may be caused by poor signal-to-noise spectra, cosmic ray artefacts in Raman spectra, or unsystematic human error, which may justify setting a limit of detection.

However, sample size, reference library size, number of blanks, and whether it is applied on sample or data set level, are just a few factors that influence the outcome of such limit determination. This is mainly due to the statistics of small numbers, which MP data scientists are faced with. Where total MP numbers are low, the ultimate probability for certain particle (pheno-)types to appear is accordingly low. In such cases, data become less robust by eliminating phenotypes below the limit of detection, hence one might conclude that 'every particle counts'. Nevertheless, an initial data quality check is recommended to identify instances of rarely occurring phenotypes, as they can lead to the detection of erroneous data entries.

3.6. Database Integration and Accessibility

An important aspect is the MPDB connection to existing international database structures. A prerequisite to comply with FAIR principles is a description of the data available in the database: the metadata. MPDB is hosted by the Leibniz Institute for Baltic Sea Research (IOW) and stored on a server under the micropoll.io-warnemuende.de domain, and the metadata are available at <http://micropoll.io-warnemuende.de/iow-metadata.xml>, accessed on 20 December 2022. The MPDB metadata are stored on a central meta-database at IOW, IOWMETA. IOWMETA transfers this meta-information to national German databases and major international databases. This ensures knowledge about the existence of MPDB and its content (Figure 5).

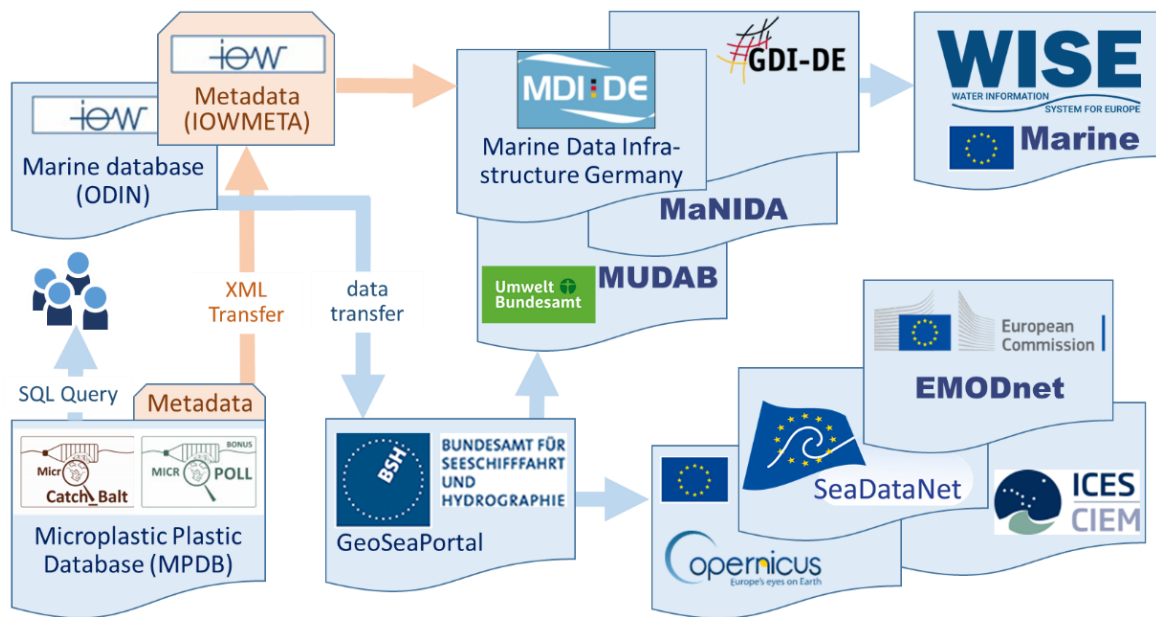


Figure 5. Integration of the microplastic database into national German and international database structures.

The second important aspect is access to the data. Presently, potential users may request MPDB access via e-mail to the main author. The potential database users are encouraged to familiarise themselves with the database documentation (Supplementary Materials).

4. Discussion

We provide the first elaborated MP-related database and aim at contributing towards a universal data repository. SQL-supporting DBMS tools provide flexible data entry and storage possibilities with many applications in various fields. The development and maintenance of such a database requires significant knowledge and resources that are hardly always available. One solution is the use of already established and tested database models. This can be achieved by replicating the DB structure from the EER diagrams (Figure 2) or directly via the provided SQL ‘create...’ statements (see Data availability statement below). The MPDB structure can be used for developing new, modified data repositories. The database re-installation and maintenance can be handled by IT departments of other research facilities or can be outsourced to such service providers. This has been done with respect to the Leibniz Institute of Polymer Research Dresden who uses and further develops an independent version of the database with a focus on microplastic in food. Alternatively, the presented MPDB version, which is hosted by the Leibniz Institute for Baltic Sea Research, can be used and further developed on request. This solution ensures that the database content is visible for other databases, through the metadata data information, and the content accessible on request.

Environmental sciences benefit from open-access data, developed and stored in structured relational databases. Taking into account that almost any sampling and analysis campaigns are costly, comparable data repositories with transparent metadata and structure are of importance for the scientific community and stakeholders.

Knowledge on emission, transport, and behaviour of microplastic in the environment presently largely results from calculations and modelling studies. Examples for the Baltic Sea are the approaches by Osinski [33], Piehl [34], or Schernewski [35,36]. A major shortcoming is the lacking result validation with microplastic field data. The MPDB and its content is a major step to overcome this weakness. The usage of structured databases ensures access to reliable data. Invalid, incomplete, or faulty data entries are detected, removed, or clarified either in the data entry phase (i.e., data with missing required fields

is not allowed in the database), or data usage (i.e., the retrieval of data, which is not related to a specific attribute will not appear in the generated view).

5. Conclusions and Outlook

The Marine Plastic Database meets all technical requirements with respect to data storage and documentation, but access to data requires expert knowledge. For end-users or stakeholders (non-researchers/scientists) the MPDB is presently not user-friendly. Several possible solutions might be examined to shape the data for easy interpretation. This process would involve two stages: (1) developing use-case scenarios for data entry/viewing/presentation, and (2) development of a front-end application or solution. Possible stakeholder use-case scenarios might be the preview of abundance of MP in their region or area of interest, prevailing MP types and their sources (media), etc. The front-end implementation solution examples would be to create an online platform for MPDB data filtering and preview, or a standalone tool (an application), developed solely for the user. To achieve this, one would require further development (web or application), extra tools and services (i.e., a server with management software). All these solutions can be explored in future uses of the MPDB. We welcome the usage of the database and are open for a collaboration aiming at an improved user-friendliness.

Supplementary Materials: The following supporting information can be downloaded at http://micropoll.io-warnemuende.de/Guideline_for_usage_of_Microplastic_Database.pdf: Guideline for Usage of Microplastic Database.

Author Contributions: Conceptualisation, database development, and writing, N.Č.; DB structure conceptualisation, support of database development, writing, visualisations, and editing, K.E., R.L., J.B., D.F., and F.F.; writing and complementary visualisation, G.S.; editing, project administration and funding acquisition, S.O. and M.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by project BONUS MICROPOLL project supported by BONUS (Article 185), funded jointly by the European Union and national funding agencies (the German Federal Ministry of Education and Research (BMBF, 03F0775A;) as well as project MicroCatch_Balt (BMBF, 03F0788A)).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available in a publicly accessible repository. The data presented in this study are openly available in a repository at Zenodo [37]. The data contain an SQL database export, which can be used to create the MPDB discussed in this article.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

DB—database; EER—enhanced entity-relationship; FAIR—findable, accessible, interoperable, and reusable; FTIR—Fourier-Transform-Infrared-Microspectroscopy, MP—microplastic; MPDB—Marine Plastic Database; PDD—particle domain data; SDD—sample domain data; SQL—structured query language.

References

1. European Commission. *Shared Environmental Information System (SEIS)*; European Commission: Brussels, Belgium, 2020; Available online: https://www.eea.europa.eu/ds_resolveuid/FWIZEG45MT (accessed on 29 December 2022).
2. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)]

3. Addamo, A.M.; Brosich, A.; Chaves Montero, M.M.; Giorgetti, A.; Hanke, G.; Molina Jack, M.E.; Vinci, M. *Marine Litter Database Lessons Learned in Compiling the First Pan-European Beach Litter Database*; Publication Office of the European Union: Luxembourg, 2018; ISBN 9789279978661. [CrossRef]
4. European Commission. *MSDF Guidance on Monitoring Marine Litter*; Publications Office of the European Union: Luxembourg, 2013; ISBN 9789279978661. Available online: <https://mcc.jrc.ec.europa.eu/documents/201702074014.pdf> (accessed on 29 December 2022).
5. Rauert, C.; Pan, Y.; Okoffo, E.D.; O'Brien, J.W.; Thomas, K.V. Extraction and Pyrolysis-GC-MS Analysis of Polyethylene in Samples with Medium to High Lipid Content. *J. Environ. Expo. Assess.* **2022**, *1*, 13. [CrossRef]
6. Fischer, M.; Scholz-Böttcher, B.M. Microplastics Analysis in Environmental Samples—Recent Pyrolysis-Gas Chromatography-Mass Spectrometry Method Improvements to Increase the Reliability of Mass-Related Data. *Anal. Methods* **2019**, *11*, 2489–2497. [CrossRef]
7. Hermabessiere, L.; Himer, C.; Boricaud, B.; Kazour, M.; Amara, R.; Cassone, A.-L.; Laurentie, M.; Paul-Pont, I.; Soudant, P.; Dehaut, A. Optimization, Performance, and Application of a Pyrolysis-GC/MS Method for the Identification of Microplastics. *Anal. Bioanal. Chem.* **2018**, *410*, 6663–6676. [CrossRef] [PubMed]
8. Primpke, S.; Lorenz, C.; Rascher-Friesenhausen, R.; Gerdt, G. An Automated Approach for Microplastics Analysis Using Focal Plane Array (FPA) FTIR Microscopy and Image Analysis. *Anal. Methods* **2017**, *9*, 1499–1511. [CrossRef]
9. Primpke, S.; Dias, P.A.; Gerdt, G. Automated Identification and Quantification of Microfibres and Microplastics. *Anal. Methods* **2019**, *11*, 2138–2147. [CrossRef]
10. Anger, P.M.; von der Esch, E.; Baumann, T.; Elsner, M.; Niessner, R.; Ivleva, N.P. Raman Microspectroscopy as a Tool for Microplastic Particle Analysis. *TrAC Trends Anal. Chem.* **2018**, *109*, 214–226. [CrossRef]
11. Brandt, J.; Bittrich, L.; Fischer, F.; Kanaki, E.; Tagg, A.; Lenz, R.; Labrenz, M.; Brandes, E.; Fischer, D.; Eichhorn, K.-J. High-Throughput Analyses of Microplastic Samples Using Fourier Transform Infrared and Raman Spectrometry. *Appl. Spectrosc.* **2020**, *74*, 1185–1197. [CrossRef]
12. Käßler, A.; Fischer, D.; Oberbeckmann, S.; Schernewski, G.; Labrenz, M.; Eichhorn, K.-J.; Voit, B. Analysis of Environmental Microplastics by Vibrational Microspectroscopy: FTIR, Raman or Both? *Anal. Bioanal. Chem.* **2016**, *408*, 8377–8391. [CrossRef] [PubMed]
13. Georgiou, A. Microplastics in the Ocean May Be Vastly Underestimated, with up to 125 Trillion Particles Floating Around, Study Says. Available online: <https://www.newsweek.com/microplastics-ocean-vastly-underestimated-125-trillion-particles-floating-around-1505404> (accessed on 29 December 2022).
14. Simon Matt Who's to Blame for Plastic Microfiber Pollution? | WIRED. Available online: <https://www.wired.com/story/whos-to-blame-for-plastic-microfiber-pollution/> (accessed on 29 December 2022).
15. Kooi, M.; Koelmans, A.A. Simplifying Microplastic via Continuous Probability Distributions for Size, Shape, and Density. *Environ. Sci. Technol. Lett.* **2019**, *6*, 551–557. [CrossRef]
16. Enders, K.; Käßler, A.; Biniash, O.; Feldens, P.; Stollberg, N.; Lange, X.; Fischer, D.; Eichhorn, K.J.; Pollehne, F.; Oberbeckmann, S.; et al. Tracing Microplastics in Aquatic Environments Based on Sediment Analogies. *Sci. Rep.* **2019**, *9*, 1–15. [CrossRef] [PubMed]
17. Bergmann, M.; Wirzberger, V.; Krumpen, T.; Lorenz, C.; Primpke, S.; Tekman, M.B.; Gerdt, G. High Quantities of Microplastic in Arctic Deep-Sea Sediments from the HAUSGARTEN Observatory. *Environ. Sci. Technol.* **2017**, *51*, 11000–11010. [CrossRef] [PubMed]
18. Enders, K.; Lenz, R.; Stedmon, C.A.; Nielsen, T.G. Abundance, Size and Polymer Composition of Marine Microplastics $\geq 10\mu\text{m}$ in the Atlantic Ocean and Their Modelled Vertical Distribution. *Mar. Pollut. Bull.* **2015**, *100*, 70–81. [CrossRef] [PubMed]
19. Frère, L.; Paul-Pont, I.; Moreau, J.; Soudant, P.; Lambert, C.; Huvet, A.; Rinnert, E. A Semi-Automated Raman Micro-Spectroscopy Method for Morphological and Chemical Characterizations of Microplastic Litter. *Mar. Pollut. Bull.* **2016**, *113*, 461–468. [CrossRef]
20. Vianello, A.; Jensen, R.L.; Liu, L.; Vollertsen, J. Simulating Human Exposure to Indoor Airborne Microplastics Using a Breathing Thermal Manikin. *Sci. Rep.* **2019**, *9*, 8670. [CrossRef]
21. Altuhafi, F.; O'Sullivan, C.; Cavarretta, I. Analysis of an Image-Based Method to Quantify the Size and Shape of Sand Particles. *J. Geotech. Geoenvironmental Eng.* **2013**, *139*, 1290–1307. [CrossRef]
22. MySQL: MySQL Installer Release Notes. Available online: <https://dev.mysql.com/doc/relnotes/mysql-installer/en/> (accessed on 29 December 2022).
23. Becker Ansgar HeidiSQL Technical Help Document. Available online: <https://www.heidisql.com/help.php#credits> (accessed on 29 December 2022).
24. Mysql-Connector-Python-PyPI. Available online: <https://pypi.org/project/mysql-connector-python/> (accessed on 29 December 2022).
25. MySQL Team: MySQL Workbench Manual. Available online: <https://dev.mysql.com/doc/workbench/en/> (accessed on 29 December 2022).
26. Wescon, I. Managing the Development of Large Software Systems. In Proceedings of the 9th International Conference on Software Engineering, Monterey, CA, USA, 30 March–2 April 1987; IEEE Computer Society Press: Washington, DC, USA, 1987; pp. 328–338.
27. Čerkasova, N.; Enders, K.; Lenz, R.; Brandt, J.; Fischer, F. Guideline for Usage of Microplastic Database. 2020. Available online: http://micropoll.io-warnemuende.de/Guideline_for_usage_of_Microplastic_Database.pdf (accessed on 29 December 2022).

28. Enders, K.; Lenz, R.; Ivar do Sul, J.A.; Tagg, A.S.; Labrenz, M. When Every Particle Matters: A QuEChERS Approach to Extract Microplastics from Environmental Samples. *MethodsX* **2020**, *7*, 100784. [[CrossRef](#)]
29. Dawson, A.L.; Santana, M.F.M.; Nelis, J.L.D.; Motti, C.A. Taking Control of Microplastics Data: A Comparison of Control and Blank Data Correction Methods. *J. Hazard. Mater.* **2023**, *443*, 130218. [[CrossRef](#)]
30. Lux, L.; Phal, Y.; Hsieh, P.-H.; Bhargava, R. On the Limit of Detection in Infrared Spectroscopic Imaging. *Appl. Spectrosc.* **2022**, *76*, 105–117. [[CrossRef](#)]
31. Klymus, K.E.; Merkes, C.M.; Allison, M.J.; Goldberg, C.S.; Helbing, C.C.; Hunter, M.E.; Jackson, C.A.; Lance, R.F.; Mangan, A.M.; Monroe, E.M.; et al. Reporting the Limits of Detection and Quantification for Environmental DNA Assays. *Environ. DNA* **2020**, *2*, 271–282. [[CrossRef](#)]
32. Saadati, N.; Abdullah, M.P.; Zakaria, Z.; Sany, S.B.T.; Rezayi, M.; Hassonizadeh, H. Limit of Detection and Limit of Quantification Development Procedures for Organochlorine Pesticides Analysis in Water and Sediment Matrices. *Chem. Cent. J.* **2013**, *7*, 63. [[CrossRef](#)] [[PubMed](#)]
33. Osinski, R.D.; Enders, K.; Gräwe, U.; Klingbeil, K.; Radtke, H. Model Uncertainties of a Storm and Their Influence on Microplastics and Sediment Transport in the Baltic Sea. *Ocean Sci.* **2020**, *16*, 1491–1507. [[CrossRef](#)]
34. Piehl, S.; Hauk, R.; Robbe, E.; Richter, B.; Kachholz, F.; Schilling, J.; Lenz, R.; Fischer, D.; Fischer, F.; Labrenz, M.; et al. Combined Approaches to Predict Microplastic Emissions Within an Urbanized Estuary (Warnow, Southwestern Baltic Sea). *Front. Environ. Sci.* **2021**, *9*, 616765. [[CrossRef](#)]
35. Schernewski, G.; Radtke, H.; Hauk, R.; Baresel, C.; Olshammar, M.; Osinski, R.; Oberbeckmann, S. Transport and Behavior of Microplastics Emissions from Urban Sources in the Baltic Sea. *Front. Environ. Sci.* **2020**, *8*, 579361. [[CrossRef](#)]
36. Schernewski, G.; Radtke, H.; Hauk, R.; Baresel, C.; Olshammar, M.; Oberbeckmann, S. Urban Microplastics Emissions: Effectiveness of Retention Measures and Consequences for the Baltic Sea. *Front. Mar. Sci.* **2021**, *8*, 594415. [[CrossRef](#)]
37. Čerkasova, N.; Enders, K.; Lenz, R.; Brandt, J.; Fischer, F. MPDB: Marine Plastics DataBase. Available online: <https://zenodo.org/record/7495007> (accessed on 30 December 2022).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.