Short documentation for the

Database on Concentrations of Hazardous Substances in the Danube River Basin

Version 1.0 Date: 2023-03-31

Authors: Steffen Kittlaus, Severin Osten (TU Wien, Institute for water quality and resource management)

skittlaus@iwag.tuwien.ac.at

1 Introduction

1.1 Target of setting up the inventory

The setup of inventories of emissions of hazardous substances into surface waters is a heavily data driven process. The most important part of this data are substance specific data which are necessary to quantify the emission loads. This can be concentrations in different environmental or technical matrices (e.g. waste water or soil), surface specific rates (e.g. atmospheric deposition rates), emissions factors for different activities (emissions per vehicle and driven kilometre) or emission loads (e.g. industrial emission reported to PRTR register). Based on such data, different types of emissions and transport models can be used to quantify the overall emissions. Such models need to be validated, which again needs measured concentration data from surface waters. To collect these data together with all necessary metadata an inventory of concentrations was created in the Danube Hazard m³c project.

The target of this activity was to collect pre-existing data from the Danube basin which were available from different sources and in different formats and combine them together with the monitoring data generated within the Danube Hazard m³c project to generate a data base as broad as possible for three applications:

1. Generate input data for emission models, especially the MoRE model (Fuchs, et al., 2017) which is mainly based on representative concentration in different pathways.
2. Supply validation data for emission models.
3. Supply harmonized data for surface water status assessment.
4. Data base for research of drivers behind the concentration patterns, which can lead to a better system understanding which can be transformed into better emission models.

1.2 Considered ideas and requirements

As this activity had a strong focus on capitalization of pre-existing knowledge, available data sources were investigated regarding their structure and the use of controlled vocabularies. These were the following data base with data available for the Danube Basin:

- Hosted by the ICPDR:
  - Transnational Monitoring Network (TMNM)
  - Joint Danube Survey Results (JDS1, JDS2, JDS3)
- Hosted by the NORMAN network:
  - Joint Danube Survey 4 Results
- Hosted by the EU
  - Reported monitoring data in the WISE framework
  - Reported emissions under the UWWT directive
  - Reported emission under the PRTR directive

Regarding the selection of the technical framework to collect the data, it was considered essential to have a system which meets the following requirements:

- Collection of data in a format, which is manageable for all contributing institutions.
- Possibility of rigorous quality control of the collected data and metadata during data import.
- Easy handling of large data amounts without any restrictions.
- Data workflow can be implemented with the available knowledge in the project team.

A data base server running the open source data base management PostgreSQL (Version 9.6) was already available at the TU Wien and is also used to run the MoRE emission model. Therefore this was chosen as technical platform. The data were migrated to PostgreSQL Version 15.2. before export to the SQL file.

2 Technical implementation of the inventory

2.1 Structure of the inventory data base

The inventory consists of many tables, which are connected by columns, which refer from one table to another. There are main tables which contain the actual data and supporting tables, which contain the allowed entries for columns with controlled vocabularies.

The main tables will be described here. They were separated into sections for different environmental and technical matrices:

- Water bodies: surface water including suspended particulate matter samples and ground water
- Waste water treatment plants: municipal and industrial, influent and effluent
- Storm water: combined sewer overflows and storm water outlets in separate sewer systems
- Atmospheric deposition
- Soil
- Sediments

As the inventory was designed to include original monitoring data wherever possible, but also data published in an aggregated form, e.g. in scientific publications and other reports, each section contains tables for single measurements as well as tables for aggregated measurements.

All tables contain the following four columns:

Column name	Data type	Obligatory	Description
id_...	serial4	x	Automatically created unique data set ID
comments	text		Comments regarding the data set

created_at	timestamptz	x	Time stamp, when the data set was created. Automatically the current system time is inserted on creation.
created_by	text	x	Creator of the data set in the data base. This information is automatically created from the user name of the importing team member on creation.

These four columns will not be listed in the following tables explaining the columns of the data base tables, as they are always the same.

2.1.1 General tables as basis for all sections

There are a few tables which are referenced from all sections, which contain essential information. The prefix "md_" in the table names is an abbreviation for "meta data".

2.1.1.1 Table "md_determinant"

This table lists all parameters for which information are collected in the different sections. This are mainly hazardous substances but also other parameters supporting the measurement of hazardous substances with additional information like total suspended solids concentration or water temperature.

Column name	**Data type**	**Obligatory**	**Description**
name_determinand	text	x	Identifier of the determinant created by combining the Id and the label from the WISE vocabulary[1]
parameter_name	text		Full determinant name
name_determinand_short	text		A short form of the determinant name for use in e.g. tables and figures.
abbreviation	varchar (20)		A common abbreviation of the substance used for presentation in plots, where only limited space is available. Must be unique in the whole data base.
further_names	text		List of further substance names
cas_number	varchar (12)		CAS registry number. As many substances have several CAS numbers, these named by ECHA[2] on the first place should be used.
further_cas_numbers	text		Further CAS registry numbers assigned to the determinant.

[1] https://dd.eionet.europa.eu/vocabulary/wise/ObservedProperty/view

[2] European Chemicals Agency: https://echa.europa.eu/de/information-on-chemicals

ec_number	varchar (9)		European Community number / List number assigned by ECHA
norman_susdat_id	text		ID used in the NORMAN Substance Database (NORMAN SusDat)

The data base contains further tables for classification of determinants according to different systems (md_determinant_classification, md_determinand_classification_system), for environmental quality standards (md_determinant_eqs) and for relations between determinants (md_determinant_relations: is one determinant the degradation product of another, are substances chemical similar, are determinants mainly used together etc.). These tables have a rather straight forward structure and are therefore not further documented.

2.1.1.2 Table "md_data_source"

All information in the inventory should be traceable to an original information source. In this table the data sources are listed with all necessary information.

Column name	**Data type**	**Obligatory**	**Description**
datasource_identifier	text	x	Identifier of the data source, must be unique.
full_citation	text	x	Full citation giving all the details needed to use it in a list of References.
publication_year	int4	x	Year of publication.
publication_type	text	x	Type of publication referencing to table "md_publication_type" with the controlled vocabulary: "International database", "National database", "Regional database", "Database of a research institution", "Project database", "Peer reviewed scientific publication", "Scientific publication", "Legal document"
publication_type_comments	text		Comments regarding the publication type
data_owner_contact_organisation	text	x	Data owner organisation
data_owner_name	text	x	Responsible person at the data owner organisation
data_owner_contact	text	x	Contact information of the data owner, e.g. email address.
data_owner_website	text		URL of website of the data owner or where the data are available.

data_license	text		Full data license information.
licensing_type	text	x	Type of license of the data set. Referencing to table "md_licensing_type" with the controlled vocabulary: "Public", "Partly public", "Spatially non-public", "Non-public"
licensing_comments	text		Comments regarding the license
data_supplier_name	text	x	Name of the person supplying the data set within the Danube Hazard m³c project. Referencing table "data_supplier_dhm3c" with the controlled vocabulary (project team members).

2.1.2 Water bodies

This section includes measurements for surface and groundwater including suspended particulate matter in surface waters.

2.1.2.1 Table "wb_single_measurements"

This table stores the measured values for data sets, where the disaggregated measurements are available. Each measurement of a substance concentration is one data set.

Column name	Data type	Obligatory	Description
sample_identifier	text	x	Sample identifier referring to table "wb_sample"
name_determinand	text	x	Parameter identifier referring to table "md_determinant"
observed_value	float8	x	Measured value, if the value was below LOQ or LOD the half of the respective value can be inserted.
unit_of_measure	text	x	Unit of measure, referring to table "md_water_unit_of_measure", where the controlled vocabulary is stored e.g. "mg/l", "µg/l", "ng/l", "°C", "mg/kg" …"
loq	float8		Value of the analytical limit of quantitation (LOQ)
value_below_loq	bool	x	Information if measurement was below LOQ (true/false)
lod	float8		Value of the analytical limit of detection (LOD)

value_below_lod	bool		Information if measurement was below LOD (true/false)
analysed_matrix	text	x	Information about the analysed part of the sample referring to table "md_analysed_matrix" where the controlled vocabulary is stored: "total (whole sample)", "dissolved (filtered sample)", "unknown (total or dissolved)", "solid",
analysed_matrix _comments	text		Comments regarding the analysed matrix (free text)
analysis_method	text	x	Name of the lab analysis method referring to table "md_analysis_method" where the controlled vocabulary is stored, e.g. "„LC-MS/MS", "ICP-OES", "GC-FID" …"unknown"
cen_iso_code _analytical _method	text	x	Identifier of the Norm describing the analytical method referring to table "md_analysis_method" where the controlled vocabulary is stored e.g. "DIN CEN/TS 15968", "ISO 11885", "EN ISO 9377-2-H53"…"unknown"
analysis_method _accredited	text	x	Information if the laboratory was accredited for the used method, referring to table "md_method_accredited" with the controlled vocabulary: "yes, fully" "no, but the values are of high reliability" "no" "unknown"
name_lab	text		Name of the laboratory, which executed the analysis, referring to table "md_laboratory"
lab_analysis_original_id	text		Identifier of lab analysis as given by the analysing lab.
datasource_identifier	text	x	Identifier of the data source, referring to the table "md_data_source" where all data sources are listed with details

2.1.2.2 Table "wb_sample"

The table contains the general information about each sample, e.g. the time and method of sampling.

Column name	Data type	Obligatory	Description
sample_identifier	text	x	Sample identifier as it was used in the project or publication where the data come from. As it must be unique in the table, a project or country specific prefix can be added.
igsn_pid	text		International Generic Sample Identifier (IGSN), a persistent identifier for samples, see https://www.igsn.org.
identifier_monitoring_site	text	x	Identifier_monitoring_site is referring to table "wb_monitoring_site" to connect the sample with its originating sampling site.
compartment	text	x	Information which kind of water body was sampled, referring to table "md_compartment" with the controlled vocabulary: "surface water", "ground water"
sampling_time	timestamptz	x	Time of sampling with time zone
sampling_method	text	x	The method how the sample was taken, referring to table "md_water_sampling_method" with the controlled vocabulary: "unknown", "grab sample", "large volume online solid phase extraction", "grab sample qualified", "spm sampling", "large volume sampler"
sampling_method_accredited	text	x	Accreditation information for the sampling method, referring to table "md_method_accredited" with controlled vocabulary: "unknown", "no", "no, but the values are of high reliability", "yes, fully"

river_flow_m3_per_sec	float8		River discharge during sampling in m³/s
gw_well_water_depth_m	float8		Water level below surface in the well during groundwater sampling. Negative values for artesian aquifers.
datasource_identifier	text	x	Identifier of the data source, referring to table "md_data_source" where details of all data sources are listed.

2.1.2.3 Table "wb_monitoring_site"

This table contains all met data about the monitoring sites for surface and ground water.

Column name	**Data type**	**Obligatory**	**Description**
identifier_monitoring_site	text	x	Identifier of the monitoring site as used in the original data source. As it must be unique in the table, a project or country specific prefix can be added.
identifier_scheme_monitoring_site	text	x	Schema where the identifier in the first column came from, e.g. "eionetmonitoringsitecode", "national code", "eumonitoringsitecode", "project code"
country	bpchar(2)	x	Country in which the sample was taken, coded by ISO 3166-1 alpha-2
name_monitoring_site	text		Name of the sampling site
coordinates_longitude	float8	x	Value of the longitude coordinate in the coordinate Reference system defined in "coordinates_crs_epsg_code"
coordinates_latitude	float8	x	Value of the latitude coordinate in the coordinate Reference system defined in

			“coordinates_crs_epsg_code”
coordinates_crs_epsg_code	int4	x	Coordinate Reference system of the coordinates used in the two previous columns coded by the European Petroleum Survey Group (EPSG) code, e.g 4326 for WGS84 -CRS
identifier_correlated_monitoring_site	text		Identifier of a correlated monitoring site, e.g. nearby river discharge gauge correlated to a quality monitoring station. Referring to the same table.
catchment_area_km2	float8		Area of the monitoring sites upstream catchment in km²
lmq_m3_s	float8		Long-term mean discharge at the monitoring site in m³/s
llq_m3_s	float8		Long-term annual mean base-flow discharge.
lq10_m3_s	float8		Long-term annual 10% percentile of discharge.
lhq1_m3_s	float8		Long-term annual flood discharge with a statistical recurrence rate of one year.
lhq10_m3_s	float8		Long-term annual flood discharge with a statistical recurrence rate of one decade.
lhq100_m3_s	float8		Long-term annual flood discharge with a statistical recurrence rate of one century.
mean_longterm_tss_mg_l	float8		Long-term mean concentration of total suspended solids in mg/l
water_body_identifier	text	x	Identifier of the water body the monitoring site is placed in, referring to table “wb_water_bodies”

gw_well_drained_depth_m	float8		Ground Water monitoring well: depth of water extraction
gw_well_longterm_water_depth_m	float8		Ground water monitoring well: longterm water level depth
gw_well_surrounding_land_use	varchar	x	Ground water monitoring well: main land use surrounding the well and potentially influencing the water quality. Referring to table "md_land_use" with the controlled vocabulary from the CORINE land cover classification (Kosztra, Büttner, Hazeu, & Arnold, 2019). For surface waterbody monitoring sites, "not applicable" should be selected.
datasource_identifier	text	x	Identifier of the data source, referring to table "md_data_source" where details of all data sources are listed.

2.1.2.4 Table "wb_water_bodies"

This table contains the waterbodies and their meta data.

Column name	Data type	Obligatory	Description
water_body_identifier	text	x	Identifier of the water body.
water_body_identifier_scheme	text	x	Scheme from which the identifier was taken.
water_body_category	text	x	Type of water body, referring to table "wb_water_body_category" with the controlled vocabulary:
water_body_name	text		Local name of the waterbody
country	bpchar(2)	x	Country in which the sample was taken, coded by ISO 3166-1 alpha-2

2.1.2.5 Table wb_waterbody_category

water_body_category	water_body_category_description
H1	Rivers on mountainous (high slope) areas; catchment area < 100 km2
H2	Rivers on mountainous (high slope) areas; 100 < catchment area < 1000 km2
H3	Rivers on mountainous (high slope) areas; 1000 < catchment area < 10000 km2
H4	Rivers on mountainous (high slope) areas; 10 000 km2 < catchment area
M1	Rivers on hilly (medium slope) areas; catchment area < 100 km2
M2	Rivers on hilly (medium slope) areas; 100 < catchment area < 1000 km2
M3	Rivers on hilly (medium slope) areas; 1000 < catchment area < 10000 km2
M4	Rivers on hilly (medium slope) areas; 10 000 < catchment area < 100 000 km2
M5	Rivers on hilly (medium slope) areas; 100000 km2 < catchment area
L1	Rivers on flat (low slope) areas; catchment area < 100 km2
L2	Rivers on flat (low slope) areas; 100 < catchment area < 1000 km2
L3	Rivers on flat (low slope) areas; 1000 < catchment area < 10000 km2
L4	Rivers on flat (low slope) areas; 10 000 < catchment area < 100 000 km2
L5	Rivers on flat (low slope) areas; 100 000 < catchment area < 1 000 000 km2
Unknown	Unknown
lake	Lakes
porous aquifer	Groundwaterbody - porous
carst aquifer	Groundwaterbody - carst
fissure aquifer	Groundwaterbody - fissure

2.1.2.6 Table "wb_temporal_aggregated_measurements"

This table contains data of measurements, which are representative for a period of time, either because grab samples from a time period were separately analysed but reported only as aggregated data set or because the sample was taken in a time integrating manner (composite sample).

Column name	**Data type**	**Obligatory**	**Description**
identifier_monitoring_site	text	x	This identifier is referring to table "wb_monitoring_site" to connect the

			measurements with its originating sampling site.
compartment	text	x	Information which kind of water body was sampled, referring to table “md_compartment” with the controlled vocabulary: “surface water”, “ground water”
name_determinand	text	x	Parameter identifier referring to table “md_determinant”
begin_sampling	timestamptz	x	The time when the sampling of the first sample began.
end_sampling	timestamptz	x	The time when the sampling of the last sample ended.
number_of_aggregated_values	int4		Number of aggregated measurements in case separately analysed samples were aggregated and only statistics are given. For single composite samples, this must be 1. If the number of aggregated samples is unknown, 1 should be used.
unit_of_measure	text	x	Unit of measure, referring to table “md_water_unit_of_measure”, where the controlled vocabulary is stored e.g. “mg/l“, “µg/l“, “ng/l“, “°C“, “mg/kg“ …”
highest_loq	float8		Highest level of quantitation relevant for the aggregated measurements or the composite sample.
values_below_loq_count	int4		How many of the aggregated measurements were below the level of quantitation. If none, value is 0.
highest_lod	float8		Highest level of detection relevant for the aggregated measurements.
values_below_lod_count	int4		How many of the aggregated measurements were below the level of detection. If none, value is 0. If level of detection is not known, this column is empty (NULL).
min_value	float8		Lowest value of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.

min_value_below_loq	bool		Was the minimum value of all aggregated measurements below LOQ?
mean_value	float8	x	Mean value of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
mean_value_below_loq	bool	x	Was the mean value of all aggregated measurements below LOQ?
max_value	float8		Maximum of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
max_value_below_loq	bool		Was the maximum value of all aggregated measurements below LOQ?
median_value	float8		Median of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
median_value_below_loq	bool		Was the median value of all aggregated measurements below LOQ?
standard_deviation	float8		Value of the standard deviation of all aggregated measurements.
analysed_matrix	text	x	Information about the analysed part of the sample referring to table “md_analysed_matrix” where the controlled vocabulary is stored: “total (whole sample)”, “dissolved (filtered sample)”, “unknown (total or dissolved)”, “solid”
analysed_matrix_comments	text		Comments regarding the analysed matrix (free text)
cen_iso_code_analytical_method	text	x	Identifier of the norm describing the analytical method referring to table “md_analysis_method” where the controlled vocabulary is stored e.g. DIN CEN/TS 15968, ISO 11885, EN ISO 9377-2-H53 …
name_lab	text	x	Name of the laboratory, which executed the analysis, referring to table “md_laboratory”

analysis_method	text	x	Name of the lab analysis method referring to table "md_analysis_method" where the controlled vocabulary is stored, e.g. "LC-MS/MS", "ICP-OES" …
analysis_method_accredited	text	x	Information if the laboratory was accredited for the used method, referring to table "md_method_accredited" with the controlled vocabulary: "yes, fully" "no, but the values are of high reliability" "no" "unknown"
sampling_method	text	x	The method how the sample was taken, referring to table "md_water_sampling_method" with the controlled vocabulary: "unknown", "grab sample", "large volume online solid phase extraction", "grab sample qualified", "SPM sampling", "large volume sampler"
sampling_method_accredited	text	x	Accreditation information for the sampling method, referring to table "md_method_accredited" with controlled vocabulary: "unknown", "no", "no, but the values are of high reliability", "yes, fully"
mean_river_flow_during_sampling_m3_per_sec	float8		The mean value of the discharge during sampling activity in m³/s.
mean_tss_during_sampling_mg_l	float8		The mean value of the total suspended solid concentration during sampling activity in mg/L.
mean_ec_during_sampling_ms_per_cm	float8		The mean value of the electrical conductivity during sampling activity in mS/cm.
sample_identifier_composite_sample	text		For single composite samples the sample identifier should be named here.
composite_sample_igsn_pid	text		For single composite samples the IGSN can be given here. International Generic Sample Identifier (IGSN) is a

			persistent identifier for samples, see https://www.igsn.org.
datasource_identifier	text	x	Identifier of the data source, referring to table "md_data_source" where details of all data sources are listed

2.1.2.7 Table "wb_spatiotemporal_aggregated_measurements_by_waterbody"

This table contains data of measurements, which are representative for a period of time and a water body and are reported only as aggregated data set. As no data were filled in, this table stays undocumented.

2.1.2.8 Table "wb_continous_measurements"

This table stores time series of in- or online measurements.

Column name	Data type	Obligatory	Description
identifier_monitoring_site	text	x	Identifier_monitoring_site is referring to table "wb_monitoring_site" to connect the measurement with its originating measurement site.
begin_measurement	timestamptz	x	Time when the measurement begins (UTC)
name_determinand	text	x	Parameter identifier referring to table "md_determinant" giving the measured parameter, e.g. discharge, turbidity or pH
observed_value	float8	x	Measured value, if the value was below LOQ or LOD the half of the respective value can be inserted.
unit_of_measure	text	x	Unit of measure, referring to table "md_water_unit_of_measure", where the controlled vocabulary is stored e.g. "mg/l" or "°C"
datasource_identifier	text	x	Identifier of the data source of the online measurement, referring to the table "md_data_source" where all

			data sources are listed with details
end_measurement_intervall	timestamptz		End time of the interval for which the measurement is.
analysis_method	text	x	Name of the lab analysis method referring to table "md_analysis_method" where the controlled vocabulary is stored, e.g. "HACH 3798-S sc, no T compensation" or "E+H radar sensor Micropilot FMR10"
cen_iso_code_analytical_method	text	x	Identifier of the norm describing the analytical method referring to table "md_analysis_method" where the controlled vocabulary is stored
name_lab	text	x	Name of the laboratory, which executed the analysis, referring to table "md_laboratory"

2.1.3 Waste water treatment plants (wwtp)

In the following tables concentration data and metadata from waste water treatment plants and other waste water systems are collected. Data are coming from different sampling points in waste water systems and treatment plants (inflow, outflow, sludge) and from different types of wwtps (municipal or industrial).

2.1.3.1 Table "wwtp_single_measurements"

This table stores concentration measurements, where the disaggregated measurement data are available. Each measurement of a substance concentration is one data set.

Column name	**Data type**	**Obligatory**	**Description**
wwtp_sample_identifier	text	x	Sample identifier connecting the measurement with a sample, referring to table "wwtp_sample"
name_determinand	text	x	Parameter identifier referring to table "md_determinant"
observed_value	float8	x	Measured value. If the value was below LOQ or LOD the half of the respective value can be inserted.

unit_of_measure	text	x	Unit of measure, referring to table "md_water_unit_of_measure", where the controlled vocabulary is stored e.g. "mg/l", "µg/l", "ng/l", "°C", "mg/kg" …"
loq	float8		Value of the analytical limit of quantitation (LOQ)
value_below_loq	bool	x	Information if measurement was below LOQ (true/false)
lod	float8		Value of the analytical limit of detection (LOD)
value_below_lod	bool		Information if measurement was below LOD (true/false)
analysis_method	text	x	Name of the lab analysis method referring to table "md_analysis_method" where the controlled vocabulary is stored, e.g. „LC-MS/MS", "ICP-OES", "GC-FID" …"
cen_iso_code_analytical_method	text	x	Identifier of the norm describing the analytical method referring to table "md_analysis_method" where the controlled vocabulary is stored e.g. DIN CEN/TS 15968, ISO 11885, EN ISO 9377-2-H53 …
analysis_method_accredited	text	x	Information if the laboratory was accredited for the used method, referring to table "md_method_accredited" with the controlled vocabulary: "yes, fully" "no, but the values are of high reliability" "no" "unknown"
analysed_matrix	text	x	Information about the analysed part of the sample referring to table "md_analysed_matrix" where the controlled vocabulary is stored: "total (whole sample)", "dissolved (filtered sample)",

			"unknown (total or dissolved)", "solid"
analysed_matrix_comments	text		Comments regarding the analysed matrix (free text)
datasource_identifier	text	x	Identifier of the data source, referring to table "md_data_source" where details of all data sources are listed.
name_lab	text	x	Name of the laboratory, which executed the analysis, referring to table "md_laboratory".

2.1.3.2 Table "wwtp_sample"

The table contains the information about each sample from waste water, fro which single measurements are available.

Column name	**Data type**	**Obligatory**	**Description**
wwtp_sample_identifier	text	x	Identifier as reported from the data source. Prefix (e.g. country code) can be added to keep it unique in the table.
igsn_pid	text		International Generic Sample Identifier (IGSN), a persistent identifier for samples, see https://www.igsn.org.
sampling_time	timestamptz	x	Time of sampling with time zone.
id_wwtp_dcp_text	text	x	Identifier of the wwtp discharge point from which the sample was collected, referring to table "wwtp_discharge_points". For inflow samples the attribution to the discharge point is not perfectly correct, but this makes it possible to work with the same structure

			for inflow and outflow data and no further problems were experienced.
wwtp_sampling_point	text	x	Place of the sampling in the wwtp system, referring to the table ”wwtp_sampling_point” with the controlled vocabulary: “inflow”, “outflow”, “primary sludge”, “excess sludge”
wwtp_sampling_method	text	x	The method how the sample was taken, referring to table “wwtp_sampling_method ” with the controlled vocabulary: “aggregation of grab samples”, “composite, not specified”, “flow-proportional composite”, “mean annual concentration calculated from annual loads”, “various”
wwtp_sampling_method_accredited	text		Accreditation information for the sampling method, referring to table “md_method_accredited ” with controlled vocabulary: “unknown”, “no”, “no, but the values are of high reliability”, “yes, fully”
flow_during_sampling_m3_s	float8		Value of the discharge during the sampling.
datasource_identifier	text	x	Identifier of the data source, referring to table “md_data_source” where details of all data sources are listed

2.1.3.3 Table "wwtp_temporal_aggregated_measurements"

This table contains data of measurements, which are representative for a period of time, either because grab samples from a time period were separately analysed but reported only as aggregated data set or because the sample was taken in a time integrating manner (composite sample).

Column name	Data type	Obligatory	Description
id_wwtp_dcp_text	text	x	ID of the wwtp discharge point referring to table "wwtp_discharge_points"
name_determinand	text	x	Parameter identifier referring to table "md_determinant"
begin_sampling	timestamptz	x	The time where the sampling begins (UTC)
end_sampling	timestamptz	x	The time where the sampling ends (UTC)
number_of_aggregated_values	int4		Number of aggregated values. 1 for single composite samples.
unit_of_measure	text	x	Unit of measure, referring to table "md_water_unit_of_measure", where the controlled vocabulary is stored e.g. "mg/l", "µg/l", "ng/l", "°C", "mg/kg" …"
highest_loq	float8		Highest level of quantitation relevant for the aggregated measurements.
values_below_loq_count	int4		How many of the aggregated measurements were below the level of quantitation (LOQ). If none, value is 0.
highest_lod	float8		Highest level of detection (LOD) for aggregated measurements. LOD for single composite samples.
values_below_lod_count	int4		How many of the aggregated measurements were below the LOD. If none, value is 0. If LOD is not known, this column is empty (NULL).

min_value	float8		Lowest value of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
min_value_below_loq	bool		Was the minimum value of all aggregated measurements below LOQ?
mean_value	float8	x	Mean value of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
mean_value_below_loq	bool	x	Was the mean value of all aggregated measurements below LOQ?
max_value	float8		Maximum of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
max_value_below_loq	bool		Was the maximum value of all aggregated measurements below LOQ?
median_value	float8		Median of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
median_value_below_loq	bool		Median of all aggregated measurements below the LOQ?
standard_deviation	float8		Value of the standard deviation from aggregated measurements
analysis_method	text	x	Name of the lab analysis method referring to table “md_analysis_method” where the controlled vocabulary is stored, e.g. „LC-MS/MS“, “ICP-OES“, “GC-FID“ …”
cen_iso_code_analytical_method	text	x	Identifier of the norm describing the analytical

			method referring to table “md_analysis_method” where the controlled vocabulary is stored e.g. DIN CEN/TS 15968, ISO 11885, EN ISO 9377-2-H53 …
analysis_method_accredited	text	x	Information if the laboratory was accredited for the used method, referring to table “md_method_accredited” with the controlled vocabulary: “yes, fully” “no, but the values are of high reliability” “no” “unknown”
analysed_matrix	text	x	Information about the analysed part of the sample referring to table “md_analysed_matrix” where the controlled vocabulary is stored: “total (whole sample)”, “dissolved (filtered sample)”, “unknown (total or dissolved)”, “solid”
analysed_matrix_comments	text		Comments regarding the analysed matrix (free text).
wwtp_sampling_point	text	x	Place of the sampling in the system, referring to table “wwtp_sampling_point” with the controlled vocabulary: “inflow”, “outflow”, “primary sludge”, “excess sludge”
sample_identifier_composite_sample	text		For single composite samples the sample identifier should be named here.
composite_sample_igsn_pid	text		For single composite samples the IGSN can be given here. International Generic Sample Identifier

			(IGSN) is a persistent identifier for samples, see https://www.igsn.org.
name_lab	text	x	Name of the laboratory, which executed the analysis, referring to table "md_laboratory".
wwtp_sampling_method	text	x	The method how the sample was taken, referring to table "wwtp_sampling_method" with the controlled vocabulary: "unknown", "grab sample" ...
sampling_method_accredited	text	x	Accreditation information for the sampling method, referring to table "md_method_accredited" with controlled vocabulary: "unknown", "no", "no, but the values are of high reliability", "yes, fully"
datasource_identifier	text	x	Identifier of the data source, referring to table "md_data_source" where details of all data sources are listed.

2.1.3.4 Table "wwtp"

This table contains general, technical and spatial information about the wastewater treatment plants (wwtp) and other investigated waste water systems.

Column name	Data type	Obligatory	Description
id_wwtp_text	text	x	identifier of the wwtp or other sampled systems.
identifier_scheme_wwtp	text	x	Tells the kind of identifier used above.
name_wwtp	text		Name of the waste water treatment plant or waste water sampling location.
country	bpchar(2)	x	Country in which the sample was taken, coded by ISO 3166-1 alpha-2.

wwtp_type	text	x	Type of the wwtp, referring to table “wwtp_type” with the controlled vocabulary: “municipal”, “industrial”, “mixed”, “other”
capacity_pe	int4		Design capacity of the wwtp in population equivalent (PE).
wwtp_capacity_class	text	x	Classification of the capacity in intervals of PE, referring to the table “wwtp_capacity_class” with the controlled vocabulary: “unknown” “PE < 2000”, “2000 <= PE < 5000”, “5000 <= PE < 10000”, “10000 <= PE < 100000”, “PE >= 100000”
connected_pe	int4		Real connected PE
connected_inh	int4		Real connected inhabitants (in comparison to above excluding industry).
share_catchment_area_combined_sewer	float8		Share of area drained by combined sewer system in the wwtp catchment.
treatment_primary	bool		Answer, if there is a primary treatment in this wwtp. The ontologies for the treatment was taken from the EU UWWTD.
treatment_secondary	bool		Answer, if there is a secondary treatment in this wwtp.
treatment_other	bool		Answer, if there are other treatments beside the primary and the secondary
treatment_n_removal	bool		Answer if there is a removal for nitrogen.
treatment_p_removal	bool		Answer if there is a removal for phosphorus.
treatment_chemical_precipitation	bool		Answer if there is a treatment step using chemical_precipitation
treatment_uv	bool		Answer if there is a disinfection by means of UV radiation.
treatment_chlorination	bool		Answer if chlorination is applied.
treatment_ozonation	bool		Answer if there is an advanced treatment using ozonation.
treatment_sand_filtration	bool		Answer if there is a sand filtration.
treatment_micro_filtration	bool		Answer if there is a special treatment using micro filtration.
treatment_other_advanced	bool		Answer if there is another method of advanced treatment.

treatment_other_advanced_specification	text		Description of the treatment in column “treatment_other_advanced”
datasource_identifier	text	x	Identifier of the data source, referring to table “md_data_source” where details of all data sources are listed.

2.1.3.5 Table “wwtp_discharge_points”

This table provides information about the discharge points of the wastewater treatment plants. One plant can have multiple discharge points. All samples are associated to a wwtp via their discharge point (which is correct for effluent samples, but questionable for inflow or sludge samples).

Column name	**Data type**	**Obligatory**	**Description**
id_wwtp_dcp_text	text	x	Identifier of the discharge point.
id_wwtp_text	text	x	Id of the wwtp, to which the discharge point belongs referring to table “wwtp”
identifier_scheme_wwtp_dcp	text	x	Tells the source of the discharge point identifier.
name_wwtp_discharge_point	text		Name of the discharge point.
country	bpchar(2)		Country in which the discharge point is located, coded by ISO 3166-1 alpha-2
coordinates_longitude	float8		Value of the longitude coordinate in the system defined in “coordinates_crs_epsg_code”
coordinates_latitude	float8		Value of the latitude coordinate in the system defined in “coordinates_crs_epsg_code”
coordinates_crs_epsg_code	int4		Coordinate reference system of the coordinates used in the two previous columns coded by the European Petroleum Survey Group (EPSG) code, e.g 4326 for WGS84 -CRS
nuts2021_level3_code	text		NUTS level 3 code to locate the discharge point in cases

			where coordinates are not available. Nomenclature of Territorial Units for Statistics or NUTS (French: Nomenclature des unités territoriales statistiques) is a geocode standard for referencing the administrative divisions of countries for statistical purposes. The standard, adopted in 2003, is developed and regulated by the European Union (European Comission, 2022).
wwtp_recipient_type	text	x	Recipient of wwtp effluent referencing table "wwtp_recipient_type" with the controlled vocabulary: "surface water", "ground water", "municipal sewer network", "soil", "other", "unknown"
water_body_identifier	text		Identifier of the receiving water body (if applicable and known), referring to table "wb_water_bodies".
receiving_water	text		Name of first receiving surface water (might not be a waterbody).
average_discharge_volume_m3_yr	int8		Mean value of the discharge per year in m³/a.
datasource_identifier	text	x	Identifier of the data source, referring to table "md_data_source" where details of all data sources are listed

2.1.3.6 Table "wwtp_group"

This table provides general, technical and spatial information about groups of waste water treatment plants with similar properties, so measurements only available in an aggregated form can be stored referring to this set of properties. No data were received for this data type in the project.

Column name	Data type	Obligatory	Description
id_wwtp_group_text	text	x	Identifier of the wwtp group.
name_wwtp_group	text		Name of the wwtp group.
country	bpchar(2)	x	Country in which the wwtps are located, coded by ISO 3166-1 alpha-2
number_of_wwtps	int4	x	Number of wwtps in this group.
wwtp_group_type	text	x	Type of the wwtps in the group, referring to table wwtp_type with the controlled vocabulary: "municipal", "industrial", "mixed", "other"
capacity_pe_min	int4		Minimum of the range of design capacity of the wwtps in the group, given in population equivalent.
capacity_pe_max	int4		Maximum of the range of design capacity of the wwtps in the group, given in population equivalent.
connected_pe_min	int4		cp. table wwtp.
connected_pe_max	int4		cp. table wwtp.
connected_inh_min	int4		cp. table wwtp.
connected_inh_max	int4		cp. table wwtp.
discharge_min_m3_yr	int8		cp. table wwtp.
discharge_max_m3_yr	int8		cp. table wwtp.
wwtp_type	text	x	Type of the wwtp, e.g. industrial, mixed, municipal, other, Reference to the table "wwtp_type"

treatment_primary_number	int4		How many among the group members have a primary treatment
treatment_secondary_number	int4		How many among the group members have a secondary treatment
treatment_other_number	int4		How many among the group members have another treatment
treatment_n_removal_number	int4		How many among the group members have a treatment nitrogen
treatment_p_removal_number	int4		How many among the group members have a treatment phosphorus
treatment_chemical_precipitation_number	int4		How many among the group members use a treatment with chemical_precipitation
treatment_uv_number	int4		How many among the group members use a treatment with uv
treatment_chlorination_number	int4		How many among the group members use a treatment with chlorination
treatment_ozonation_number	int4		How many among the group members use a treatment with ozonation
treatment_sand_filtration_number	int4		How many among the group members use a treatment with sand filtration
treatment_micro_filtration_number	int4		How many among the group members use a treatment with micro filtration
datasource_identifier	text	x	Identifier of the data source, referring to table

			"md_data_source" where details of all data sources are listed

2.1.3.7 Table "wwtp_spatiotemporal_aggregated_measurements"

This table contains data of measurements, which are representative for a period of time, either because grab samples from a time period were separately analysed but reported only as aggregated data set or because the sample was taken in a time integrating manner (composite sample). In addition to "wwtp_temporal_aggregated_measuremnts", this information is merged not only of one wwtp, but from many wwtps. The results are therefore representative for a whole group of wwtp during a specific time period. No data were received for this data type in the project.

Column name	**Data type**	**Obligatory**	**Description**
id_wwtp_group	int4	x	Id_wwtp_group_text referring to table "wwtp_group"
begin_sampling	timestamptz	x	The time where the sampling begins (UTC)
end_sampling	timestamptz	x	The time where the sampling ends (UTC)
name_determinand	text	x	parameter identifier referring to table "md_determinand"
number_of_aggregated_values	int4		Number of aggregated determinads
unit_of_measure	text	x	Unit of measure, referring to table "md_water_unit_of_measure", where the controlled vocabulary is stored e.g. "mg/l", "µg/l", "ng/l", "°C", "mg/kg" …"
highest_loq	double precision		Highest level of quantitation relevant for the aggregated measurements.
values_below_loq_count	int4	x	How many of the aggregated measurements were below the level of quantitation. If none, value is 0.
highest_lod	float8		Highest level of determination for aggregated measurements.

values_below_lod_count	int4		How many of the aggregated measurements were below the level of detection. If none, value is 0. If level of detection is not known, this column is empty (null).
min_value	float8		Lowest value of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
min_value_below_loq	bool		Was the minimum value of all aggregated measurements below LOQ?
mean_value	float8	x	Mean value of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
mean_val_below_loq	bool	x	Mean of all aggregated measurements below the level of quantification.
max_value	float8		Maximum of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
max_value_below_loq	bool		Was the maximum value of all aggregated measurements below LOQ?
median_value	float8		Median of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
median_value_below_loq	bool		Median of all aggregated measurements below the level of quantification?
standard_deviation	float8		Value of the standard deviation including all aggregated measurements.
analysis_method	text	x	Name of the lab analysis method referring to table “md_analysis_method” where the controlled vocabulary is stored, e.g. „LC-MS/MS“, “ICP-OES“, “GC-FID“ …”

cen_iso_code_analytical_method	text	x	Identifier of the Norm describing the analytical method referring to table "md_analysis_method" where the controlled vocabulary is stored e.g. DIN CEN/TS 15968, ISO 11885, EN ISO 9377-2-H53 …
analysis_method_accredited	text	x	Information if the laboratory was accredited for the used method, referring to table "md_method_accredited" with the controlled vocabulary: "yes, fully" "no, but the values are of high reliability" "no" "unknown"
name_lab	text	x	Name of the laboratory, which executed the analysis, referring to table "md_laboratory"
analysed_matrix	text	x	Information about the analysed part of the sample referring to table "md_analysed_matrix" where the controlled vocabulary is stored: "total (whole sample)", "dissolved (filtered sample)", "unknown (total or dissolved)", "solid"
analysed_matrix_comments	text		Comments regarding the analysed matrix (free text)
wwtp_sampling_method	text	x	The method how the sample was taken, referring to table "wwtp_sampling_method" with the controlled vocabulary: "aggregation of grab samples", "passive sample", "various", "unknown", "standard", "mean annual

			concentration calculated from annual loads", "flow-proportional composite", "statistical aggregation of measurements in grab samples", "composite, not specified", "composite, 4-hours"
sampling_method_accredited	text	x	Accreditation information for the sampling method, referring to table "md_method_accredited" with controlled vocabulary: "unknown", "no", "no, but the values are of high reliability", "yes, fully"
wwtp_sampling_point	text	x	Place of the sampling, referring to table "wwtp_sampling_point", where the controlled vocabulary is stored: "excess sludge", "inflow", "outflow", "primary sludge"
datasource_identifier	text	x	Identifier of the data source, referring to table "md_data_source" where details of all data sources are listed.

2.1.4 Storm water

This section contains measurements from storm water outflows, both from combined sewers and from separate sewer systems.

2.1.4.1 Table "stw_single_measurements"

This table stores the measured values for data sets, where the disaggregated measurements are available. Each measurement of a substance concentration is one data set.

Column name	**Data type**	**Obligatory**	**Description**
id_stw_sample_text	text	x	Id_stw_sample_text referring to table "stw_sample"
name_determinand	text	x	Parameter identifier referring to table "md_determinant"

unit_of_measure	text	x	Unit of measure, referring to table "md_water_unit_of_measure", where the controlled vocabulary is stored e.g. "mg/l", "µg/l", "ng/l", "°C", "mg/kg" …"
observed_value	float8	x	Measured value, if the value was below LOQ or LOD the half of the respective value can be inserted.
loq	float8		Value of the analytical limit of quantitation (LOQ)
value_below_loq	bool	x	Information if measurement was below LOQ (true/false)
lod	float8		Value of the analytical limit of detection (LOD)
value_below_lod	bool		Information if measurement was below LOD (true/false)
analysis_method	text	x	Name of the lab analysis method referring to table "md_analysis_method" where the controlled vocabulary is stored, e.g. "LC-MS/MS", "ICP-OES", "GC-FID" …"
cen_iso_code_analytical_method	text	x	Identifier of the norm describing the analytical method referring to table "md_analysis_method" where the controlled vocabulary is stored e.g. DIN CEN/TS 15968, ISO 11885, EN ISO 9377-2-H53 …
name_lab	text		Name of the laboratory, which executed the analysis, referring to table "md_laboratory"
analysis_method_accredited	text	x	Information if the laboratory was accredited for the used method, referring to table "md_method_accredited" with the controlled vocabulary: "yes, fully"

			“no, but the values are of high reliability” “no” “unknown”
analysed_matrix	text		Information about the analysed part of the sample referring to table “md_analysed_matrix” where the controlled vocabulary is stored: “total (whole sample)”, “dissolved (filtered sample)”, “unknown (total or dissolved)”, “solid”
analysed_matrix_comments	text		Comments regarding the analysed matrix (free text)
datasource_identifier	text	x	Identifier of the data source, referring to table “md_data_source” where details of all data sources are listed

2.1.4.2 Table “stw_sample”

The table contains the general information about each sample, e.g. the time and method of sampling.

Column name	Data type	Obligatory	Description
id_stw_sample_text	text	x	
igsn_pid	text		International Generic Sample Identifier (IGSN), a persistent identifier for samples, see https://www.igsn.org.
id_stw_monitoring_site_text	text	x	Id_stw_monitoring_site_text referring to table “stw_monitoring_site”
begin_sampling	timestamptz		The time where the sampling begins (UTC)
end_sampling	timestamptz		The time where the sampling ends (UTC)
stw_type_sampling_point	text	x	Place of sampling in the system, referring to table “stw_type_sampling_point” where the controlled vocabulary is stored: “outlet of

			a sewer (without treatment)", "after sedimentation tank", "after soil filters", "other", "unknown"
stw_sampling_method	text	x	The method how the sample was taken, referring to table "stw_sampling_method" with the controlled vocabulary: "grab sample", "time proportional sample", "other", "unknown", "flow proportional sample"
sampling_method_accredited	text	x	Accreditation information for the sampling method, referring to table "md_method_accredited" with controlled vocabulary: "unknown", "no", "no, but the values are of high reliability", "yes, fully"
discharge_during_sampling_m3_s	float8		Mean discharge during sampling in m³/s.
composite_no_fillings	float8		How many sub samples were combined into one composite sample.
composite_no_overflows	float8		Number of overflow events combined into one composite sample.
composite_share_of_vol_sampled_percent	float8		Percent of event volume which was successfully sampled.
datasource_identifier	text	x	Identifier of the data source, referring to table "md_data_source" where details of all data sources are listed

2.1.4.3 Table "stw_monitoring_site"

This table contains all general information about storm water monitoring sites.

Column name	**Data type**	**Obligatory**	**Description**
id_stw_monitoring_site_text	text	x	Identifier of the monitoring site. Must be unique.

identifier_scheme_stw_monitoring_site	text		Where does the identifier come from which was used in the column before?
country	bpchar(2)	x	Country in which the sampling site is located coded by ISO 3166-1 alpha-2
name_stw_monitoring_site	text		Local name of the monitoring site
coordinates_longitude	float8		Value of the longitude coordinate of the system defined in "coordinates_crs_epsg_code"
coordinates_latitude	float8		Value of the latitude coordinate of the system defined in "coordinates_crs_epsg_code"
coordinates_crs_epsg_code	int4		Coordinate Reference system of the coordinates used in the two previous columns coded by the European Petroleum Survey Group (EPSG) code, e.g 4326 for WGS84 -CRS
stw_type_sewer_system	text	x	Type of the sewer system, referring to table "stw_type_sewer_system" with controlled vocabulary: "combined", "separated", "unknown"
catchment_area_tot_km2	float8		Size of the total catchment area of the sewer outlet in km².
catchment_area_imp_km2	float8		Size of the impervious catchment area of the sewer outlet in km².
catchment_area_imp_conn_km2	float8		Size of the impervious and connected catchment area of the sewer outlet in km².
discharge_m3_yr	float8		Typical discharge at the monitoring site in m³/year.

related_precipitation_gauge	text		Related precipitation gauge, referring to table “precipitation_gauge”.
catchment_area_conn_km2	float8		Directly connected catchment area (without upstream overflow facilities) in km².
catchment_area_conn_traffic_km2	float8		Traffic areas in the directly connected catchment area in km²
catchment_area_conn_industry_km2	float8		Industrial and commercial areas in the directly connected catchment area in km².
catchment_inh	int4		population in the catchment in inhabitants.
facility_v_storage_m3	float8		Storage volume of the CSO/SSO facility in m³.
facility_q_throttle_l_s	float8		Throttled flow to the wastewater treatment plant in l/s.
catchment_mean_prec_mm_a	float8		Average annual precipitation in mm/year
datasource_identifier	text	x	Identifier of the data source, referring to table “md_data_source” where details of all data sources are listed

2.1.4.4 Table “stw_temporal_aggregated_measurements”

This table contains data of measurements, which are representative for a period of time, either because grab samples from a time period were separately analysed but reported only as aggregated data set or because the sample was taken in a time integrating manner (composite sample). No data of this type was received in the project.

Column name	**Data type**	**Obligatory**	**Description**
id_stw_monitoring_site_text	text	x	Identifier of the sampling site referring to table “stw_monitoring_site”
stw_type_sampling_point	text	x	Place of sampling in the system, referring to table “stw_type_sampling_point”

			where the controlled vocabulary is stored: “outlet of a sewer (without treatment)”, “after sedimentation tank”, “after soil filters”, “other”, “unknown”
sample_identifier_composite_sample	text		Sample identifier in case the entry is for a single composite sample.
composite_sample_igsn_pid	text		For single composite samples the IGSN can be given here. International Generic Sample Identifier (IGSN) is a persistent identifier for samples, see https://www.igsn.org.
name_determinand	text	x	Parameter identifier referring to table “md_determinant”
begin_sampling	timestamptz	x	The time where the sampling begins (UTC)
end_sampling	timestamptz	x	The time where the sampling ends (UTC)
unit_of_measure	text	x	Unit of measure, referring to table “md_water_unit_of_measure”, where the controlled vocabulary is stored e.g. “mg/l“, “µg/l“, “ng/l“, “°C“, “mg/kg“ …
analysis_method	text	x	Name of the lab analysis method referring to table “md_analysis_method” where the controlled vocabulary is stored, e.g. „LC-MS/MS“, “ICP-OES“, “GC-FID“ …
cen_iso_code_analytical_method	text	x	Identifier of the norm describing the analytical method referring to table “md_analysis_method” where the controlled vocabulary is stored e.g. DIN

			CEN/TS 15968, ISO 11885, EN ISO 9377-2-H53 …
analysis_method_accredited	text	x	Information if the laboratory was accredited for the used method, referring to table “md_method_accredited” with the controlled vocabulary: “yes, fully” “no, but the values are of high reliability” “no” “unknown”
number_of_aggregated_values	int4		Number of aggregated measurements.
highest_loq	float8		Highest level of quantitation relevant for the aggregated measurements.
values_below_loq_count	int4		How many of the aggregated measurements were below the level of quantitation? If none, value is 0.
highest_lod	float8		Highest level of determination for aggregated measurements
values_below_lod_count	int4		How many of the aggregated measurements were below the level of detection? If none, value is 0. If level of detection is not known, this column is empty (null).
min_value	float8		Lowest value of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
min_value_below_loq	bool		Lowest value of all aggregated measurements below the level of quantification
mean_value	float8	x	Mean value of all aggregated measurements. If value below LOQ or LOD,

			preferably report half value of LOQ/LOD here.
mean_value_below_loq	bool	x	Was the mean value of all aggregated measurements below LOQ?
max_value	float8		Maximum of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
max_value_below_loq	bool		Was the maximum value of all aggregated measurements below LOQ?
median_value	float8		Median of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
median_value_below_loq	bool		Median of all aggregated measurements below the level of quantification?
standard_deviation	float8		Value of the standard deviation of all aggregated measurements.
analysed_matrix	text	x	Information about the analysed part of the sample referring to table “md_analysed_matrix” where the controlled vocabulary is stored: “total (whole sample)”, “dissolved (filtered sample)”, “unknown (total or dissolved)”, “solid”
analysed_matrix_comments	text		Comments regarding the analysed matrix (free text)
stw_sampling_method	text	x	The method how the sample was taken, referring to table “stw_sampling_method” with the controlled vocabulary: “grab sample”, “time proportional sample”,

			"other", "unknown", "flow proportional sample"
sampling_method_accredited	text	x	Accreditation information for the sampling method, referring to table "md_method_accredited" with controlled vocabulary: "unknown", "no", "no, but the values are of high reliability", "yes, fully"
name_lab	text	x	Name of the laboratory, which executed the analysis, referring to table "md_laboratory".
datasource_identifier	text	x	Identifier of the data source, referring to table "md_data_source" where details of all data sources are listed

2.1.5 Atmospheric deposition

Atmospheric deposition is the process of transferring substances from the atmosphere, where they might be available as gas, as aerosol or as suspended dust, onto surfaces during dry weather (dry deposition) or during rain and snowfall (wet deposition).

Dry deposition is caused by condensation and sedimentation of atmospheric matter, while wet deposition is additionally caused by a wash-out effects.

Atmospheric deposition can not be easily derived from air concentration measurements, therefore it requires its own monitoring set up. Nevertheless, bulk deposition (dry + wet deposition) is often reported as concentration, but referring to precipitation volume (or even dry matter in the precipitation), not to air volume. It is important to have the duration of sampling, the area of the sampler surface and the sample volume available to calculate deposition rates (mass/(area·time)), the finally needed information.

2.1.5.1 Table "ad_single_measurements"

This table stores the measured values for data sets, where the disaggregated measurements are available. Each measurement of a substance concentration or deposition rate is one data set.

Column name	Data type	Obligatory	Description

id_ad_sample_text	text	x	Identifier of the sample, referring to table “ad_sample”
name_determinand	text	x	Parameter identifier referring to table “md_determinant”
observed_value	float8	x	Measured value, if the value was below LOQ or LOD the half of the respective value can be inserted.
ad_unit_of_measure	text	x	Unit of measure, referring to table “ad_unit_of_measure”, where the controlled vocabulary is stored e.g. “µg/l”, “µg/m²/d“, “mg/kg DM“
loq	float8		Value of the analytical limit of quantitation (LOQ)
value_below_loq	bool	x	Information if measurement was below LOQ?
lod	float8		Value of the analytical limit of detection (LOD)
value_below_lod	bool		Information if measurement was below LOD?
name_lab	text	x	Name of the laboratory, which executed the analysis, referring to table “md_laboratory”.
ad_analysed_matrix	text		Information about the analysed part of the sample referring to table “md_analysed_matrix” where the controlled vocabulary is stored: “total (whole sample)”, “dissolved (filtered sample)”, “unknown (total or dissolved)”, “solid”

cen_iso_code_analytical_method	text	x	Identifier of the norm describing the analytical method referring to table "md_analysis_method" where the controlled vocabulary is stored e.g. DIN CEN/TS 15968, ISO 11885, EN ISO 9377-2-H53 …
analysis_method	text	x	Name of the lab analysis method referring to table "md_analysis_method" where the controlled vocabulary is stored, e.g. „LC-MS/MS", "ICP-OES", "GC-FID" …"
analysis_method_accredited	text	x	Information if the laboratory was accredited for the used method, referring to table "md_method_accredited" with the controlled vocabulary: "yes, fully" "no, but the values are of high reliability" "no" "unknown"
datasource_identifier	text	x	Identifier of the data source, referring to table "md_data_source" where details of all data sources are listed

2.1.5.2 Table "ad sample"

The table contains the general information about each sample, e.g. the time and method of sampling.

Column name	Data type	Obligatory	Description
id_ad_sample_text	text	x	Identifier of the sample.
igsn_pid	text		International Generic Sample Identifier (IGSN), a

			persistent identifier for samples, see https://www.igsn.org.
id_ad_sampling_site_text	text	x	Identifier of the sampling site, referring to table “ad_sampling_site”
begin_sampling	timestamptz	x	The time when the sampling begins with time zone
end_sampling	timestamptz	x	The time when the sampling ends with time zone
sampling_method_accredited	text	x	Accreditation information for the sampling method, referring to table “method_accredited” with controlled vocabulary: “unknown”, “no”, “no, but the values are of high reliability”, “yes, fully”
ad_sample_type	text	x	Type of sample, referring to table “ad_sample_type” with the controlled vocabulary: “wet only”, “dry only”, “bulk”
dry_content_concentration_mg_l	float8		Solids concentration in the sample in mg/l.
precipitation_sample_mm	float8		Value of the precipitation in the sample in mm.
annual_precipitation_sampling_year_mm	float8		Annual precipitation in the year the sample was taken in mm.
surface_area_sampler_cm2	float8		Projected surface area of the sampling device in cm².
sample_volume_ml	float8		Collected sample amount in mL.
measured_precipitation_during_sampling_mm	float8		Value of the precipitation during the sampling period from an independent precipitation gauge.

datasource_identifier	text	x	Identifier of the data source, referring to table “md_data_source” where details of all data sources are listed.

2.1.5.3 Table “ad_sampling_site”

This table contains all general information about atmospheric deposition sampling sites.

Column name	**Data type**	**Obligatory**	**Description**
id_ad_sampling_site_text	text	x	Identifier of the sampling site, must be unique.
country	bpchar(2)	x	Country in which the sample was taken, coded by ISO 3166-1 alpha-2
name_sampling_site	text		Local name of the sampling site.
coordinates_longitude	float8		Value of the longitude coordinate of the system defined in “coordinates_crs_epsg_code”
coordinates_latitude	float8		Value of the latitude coordinate of the system defined in “coordinates_crs_epsg_code”
coordinates_crs_epsg_code	float8		Coordinate Reference system of the coordinates used in the two previous columns coded by the European Petroleum Survey Group (EPSG) code, e.g 4326 for WGS84 -CRS
ad_site_type	text	x	Classification of the sampling site regarding potential main pollution sources, referring to table “ad_sample_type” with the controlled vocabulary:

			“background”, “urban”, “industrial”, “unknown”, “mixed”, “agriculture”, “rural”
longterm_mean_annual_precipitation_mm	float8		Longterm mean annual precipitation at the sampling site in mm/year.
related_precipitation_gauge	text		Identifier of a correlated precipitation monitoring site, referring to table “precipitation_gauge”.
nuts1_unit	text		NUTS level 1 code. In cases where the exact location of the station cannot be disclosed, the NUTS units can be used to give rough localization. Nomenclature of Territorial Units for Statistics or NUTS (French: Nomenclature des unités territoriales statistiques) is a geocode standard for referencing the administrative divisions of countries for statistical purposes. The standard, adopted in 2003, is developed and regulated by the European Union (European Comission, 2022).
nuts2_unit	text		NUTS level 2 code. See above.
nuts3_unit	text		NUTS level 3 code. See above.
datasource_identifier	text	x	Identifier of the data source, referring to table “md_data_source” where details of all data sources are listed

2.1.5.4 Table "ad_temporal_aggregated_measurements"

This table contains data of measurements, which are representative for a period of time, because samples from several time periods were separately analysed but reported only as aggregated data set. No data of this type were received in the project.

Column name	Data type	Obligatory	Description
id_ad_sampling_site_text	text	x	Identifier of the sampling site, referring to table "ad_sampling_site"
name_determinand	text	x	Parameter identifier referring to table "md_determinant"
begin_sampling	timestamptz	x	The time where the sampling begins (UTC)
end_sampling	timestamptz	x	The time where the sampling ends (UTC)
number_of_aggregated_values	int4		Number of aggregated measurements.
ad_unit_of_measure	text	x	Unit of measure, referring to table "ad_unit_of_measure", where the controlled vocabulary is stored e.g. "µg/l", "µg/m²/d", "mg/kg DM"
highest_loq	float8		Highest LOQ relevant in the aggregated data.
values_below_loq_count	int4		How many of the aggregated measurements were below the level of quantitation? If none, value is 0.
highest_lod	float8		Highest level of determination in the aggregated measurements.
values_below_lod_count	int4		How many of the aggregated measurements were below the level of detection? If none,

			value is 0. If level of detection is not known, this column is empty (null).
min_value	float8		Lowest value of all aggregated measurements. If value below loq or lod, preferably report half value of loq/lod here.
min_value_below_loq	bool		Was the minimum value of all aggregated measurements below LOQ?
mean_value	float8	x	Mean value of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
mean_value_below_loq		x	Was the mean value of all aggregated measurements below LOQ?
max_value	float8		Maximum of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
max_value_below_loq	bool		Was the maximum value of all aggregated measurements below LOQ?
median_value	float8		Median of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
median_value_below_loq	bool		Median of all aggregated measurements below

			the level of quantification
standard_deviation	float8		Value of the standard deviation derived from the aggregated measurements.
analysis_method	text	x	Name of the lab analysis method referring to table "md_analysis_method" where the controlled vocabulary is stored, e.g. „LC-MS/MS", "ICP-OES", "GC-FID" …"
cen_iso_code_analytical_method	text	x	Identifier of the norm describing the analytical method referring to table "md_analysis_method" where the controlled vocabulary is stored e.g. DIN CEN/TS 15968, ISO 11885, EN ISO 9377-2-H53 …
analysis_method_accredited	text	x	Information if the laboratory was accredited for the used method, referring to table "md_method_accredited" with the controlled vocabulary: "yes, fully" "no, but the values are of high reliability" "no" "unknown"
ad_sample_type	text	x	Type of sample, referring to table "ad_sample_type" with the controlled vocabulary:

			“wet only”, “dry only”, “bulk”
precipitation_sum_samples_mm	float8		Sum of precipitation during collection of all samples.
annual_precipitation_sampling_year_mm	float8		Annual precipitation in the year of sampling in mm.
sample_identifier_composite_sample	text		For single composite samples the sample identifier should be named here.
name_lab	text	x	Name of the laboratory, which executed the analysis, referring to table “md_laboratory”
Ad_analysed_matrix	text		Information about the analysed part of the sample referring to table “md_analysed_matrix” where the controlled vocabulary is stored: “total (whole sample)”, “dissolved (filtered sample)”, “unknown (total or dissolved)”, “solid”
datasource_identifier	text	x	Identifier of the data source, referring to table “md_data_source” where details of all data sources are listed

2.1.6 Soil

Soil can contribute to surface water pollution via soil erosion. Therefore concentration in top soils were collected in the following tables.

2.1.6.1 Table “soil_single_measurements”

This table stores the measured values for data sets, where the disaggregated measurements are available. Each measurement of a substance concentration is one data set.

Column name	Data type	obligatory	Description
id_soil_sample_text	text	x	Id_soil_sample_text referring to table “soil_sample”
name_determinand	text	x	Parameter identifier referring to table “md_determinant”
extraction_method	text		Method used to extract the soil, referring to table “md_solids_extraction_method”
unit_of_measure	text	x	Unit of measure, referring to table “md_solids_unit_of_measure”, where the controlled vocabulary is stored: “mg/kg WW”, “mg/kg DM“, “µg/kg DM“, “-“
observed_value	float8	x	Measured value, if the value was below LOQ or LOD the half of the respective value can be inserted.
analysis_method	text	x	Name of the lab analysis method referring to table “md_analysis_method” where the controlled vocabulary is stored, e.g. „LC-MS/MS“, “ICP-OES“, “GC-FID“ …”
cen_iso_code_analytical_method	text	x	Identifier of the Norm describing the analytical method referring to table “md_analysis_method” where the controlled vocabulary is stored e.g. DIN CEN/TS 15968, ISO 11885, EN ISO 9377-2-H53 …
analysis_method_accredited	text	x	Information if the laboratory was accredited for the used method, referring to table “md_method_accredited” with the controlled vocabulary: “yes, fully” “no, but the values are of high reliability” “no” “unknown”
loq	float8		Value of the analytical limit of quantitation (LOQ)

value_below_loq	bool	x	Information if measurement was below LOQ
lod	float8		Value of the analytical limit of detection (LOD)
value_below_lod	bool		Information if measurement was below LOD
name_lab	text	x	Name of the laboratory, which executed the analysis, referring to table “md_laboratory”, where all known labs are listed.
datasource_identifier	text	x	Identifier of the data source, referring to table “md_data_source” where details of all data sources are listed.

2.1.6.2 Table “soil_sample”

The table contains the general information about each soil sample, e.g. the time and method of sampling.

Column name	**Data type**	**Obligatory**	**Description**
id_soil_sample_text	text		Identifier of the soil sample.
igsn_pid	text		International Generic Sample Identifier (IGSN), a persistent identifier for samples, see https://www.igsn.org.
sampling_date			Date on which the sample was taken
soil_sampling_fraction	text	x	Part of the soil column, which was sampled, referring to table “soil_sample_fraction” with the controlled vocabulary: “Top soil 0-5 cm”, “Top soil 0-10 cm”, “Top soil 0-20 cm”, “Top soil 0-30 cm”, “Top soil 0-50 cm”, “Sub soil”, “Top soil”, “Humus cover”, “unknown”, “Top soil 0-

			15 cm", "Top soil 0-40 cm"
soil_sampling_method	text	x	The method how the sample was taken, referring to table "soil_sampling_method" with the controlled vocabulary: "from profile", "cutting frame (20 x 20 cm)", "core or sleeve-type borer (split-tube sampler)", "unknown", "grass plot sampler", "gouge auger (Typ Pürckhauer)", "soil ring", "cutting square or shovel (25 x 25 cm)", „hand driven auger (Typ Edelman/Schärpe)", „drilling (HU-NEBIH-TIM_standard)"
soil_sampling_method_accredited	text	x	Accreditation information for the sampling method, referring to table "md_method_accredited" with controlled vocabulary: "unknown", "no", "no, but the values are of high reliability", "yes, fully"
dray_contents_sample	float 8		Share of dry matter from the total sample weight.
organic_content_sample_mgc_per_kg_dry	float 8		Organic content in the sample given in mg carbon/kg.
datasource_identifier	text	x	Identifier of the data source, referring to table "md_data_source" where details of all data sources are listed.

2.1.6.3 Table "soil_sampling_site"

This table contains all general information about soil sampling sites.

Column name	Data type	Obligatory	Description
id_soil_sampling_site_text	text	x	Identifier of the soil sampling site.
identifier_scheme_sampling_site	text		Source of the Identifier used in the previous column.
country	bpchar(2)	x	Country in which the sample was taken, coded by ISO 3166-1 alpha-2
name_soil_sampling_site	text		Name of the sampling site.
coordinates_longitude	float8	x	Value of the longitude coordinate of the system defined in "coordinates_crs_epsg_code"
coordinates_latitude	float8	x	Value of the latitude coordinate of the System defined in "coordinates_crs_epsg_code"
coordinates_crs_epsg_code	int4	x	Coordinate Reference system of the coordinates used in the two previous columns coded by the European Petroleum Survey Group (EPSG) code, e.g 4326 for WGS84 -CRS
soil_texture	text		Texture of the soil at the sampling site, referring to table "soil_texture" with the controlled vocabulary based on the FAO soil texture classification with 12 classes (Food and Agriculture Organization of the United Nations (FAO), 2006).
soil_genetic_type	text		Describes the soil genetic type, e.g. rendzina, brown forest soil, young raw casting soils and is referring to table "soil_genetic_type", where the vocabulary is listed. Unfortunately, it was not harmonized yet, as genetic soil classification systems

			differ in the different countries quite a lot.
land_use	text		Usage/coverage of the land at the soil sampling site, referring to table "md_land_use" with the controlled vocabulary from the CLC classification.
datasource_identifier	text	x	Identifier of the data source, referring to table "md_data_source" where details of all data sources are listed

2.1.6.4 Table "soil_sample_belonging_to_sample_site"

As one soil sample can be composed from sub samples from different sampling sites and multiple samples can be analysed from on sampling site (e.g. different horizons), the relation between soil samples and soil sampling sites is many to many. Therefore, the references are done in this extra table.

Column name	**Data type**	**Obligatory**	**Description**
id_soil_sample_text	text	x	Identifier of the soil sample
id_soil_sampling_site_text	text	x	Identifier of the corresponding sampling site

2.1.6.5 Table "soil_spatial_aggregated_measurements"

This table contains data of measurements, which are given aggregated over a spatial domain. No data of this type were received in the project. Column descriptions preliminary.

Column name	**Data type**	**Obligatory**	**Description**
identifier_soil_sp_agg_val	text	x	Identifier of the data set.
country	bpchar(2)	x	Country in which the sample was taken, coded by ISO 3166-1 alpha-2
name_determinand	text		Parameter identifier referring to table "determinand"
begin_sampling	timestamptz		The time where the sampling begins (UTC)
end_sampling	timestamptz	x	The time where the sampling ends (UTC)

unit_of_measure	text		unit of measure, referring to table “solids_unit_of_measure”, where the controlled vocabulary is stored e.g. “mg/l“, “µg/l“, “ng/l“, “°C“, “mg/kg“ …”
highest_loq	float8		Highest level of quantitation relevant for the aggregated measurements
values_below_loq_count	int4		How many of the aggregated measurements were below the level of quantitation. If none, value is 0.
highest_lod	float8		Highest level of determination for aggregated determinands
values_below_lod_count	int4		How many of the aggregated measurements were below the level of detection. If none, value is 0. If level of detection is not known, this column is empty (null).
min_value	float8		Lowest value of all aggregated measurements. If value below loq or lod, preferably report half value of loq/lod here.
min_value_below_loq	bool		Was the minimum value of all aggregated measurements below LOQ?
mean_value	float8	x	Mean value of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
mean_value_below_loq	bool	x	Was the mean value of all aggregated measurements below LOQ?
max_value	float8		Maximum of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
max_value_below_loq	bool		Was the maximum value of all aggregated measurements below LOQ?
median_value	float8		Median of all aggregated measurements. If value below LOQ or LOD, preferably report half value of LOQ/LOD here.
median_value_below_loq	bool		Median of all aggregated determinands below the level of quantification
standard_deviation	float8		Value of the standard deviation including all aggregated determinands

soil_sampling_method	text		The method how the sample was taken, referring to table “soil_sampling_method” with the controlled vocabulary: “unknown”, “grab sample”, “large volume online solid phase extraction”, “grab sample qualified”, “SPM sampling”, “large volume sampler”
soil_sample_fration	text		
extraction_method	text		Method used to extract the soil, Reference to the table “extraction_method”
analysis_method	text	x	Name of the lab analysis method referring to table “analysis method” where the controlled vocabulary is stored, e.g. „LC-MS/MS“, “ICP-OES“, “GC-FID“ …”
cen_iso_code_analytical_method	text	x	Identifier of the Norm describing the analytical method referring to table “analysis method” where the controlled vocabulary is stored e.g. DIN CEN/TS 15968, ISO 11885, EN ISO 9377-2-H53 …
analysis_method_accredited	text	x	Information if the laboratory was accredited for the used method, referring to table “method_accredited” with the controlled vocabulary: “yes, fully” “no, but the values are of high reliability” “no” “unknown”
soil_texture	text		Describes the txture of the soil, e.g. loam, silty clay loam, sandy loam…, Reference to the table “soil_texture”
soil_genetic_type	text		Describes the genetic type, e.g. rendzina, brown forest soil, young raw casting soils
land_use	text		Usage of the surrounding land, Reference to table “land_use”
datasource_identifier	text	x	Identifier of the data source, referring to table “data_source” where details of all data sources are listed
name_lab	text	x	Name of the laboratory, which executed the analysis, Refering to table “laboratory”

2.1.7 Sediments

Tables for sediments were drafted, but no data for sediments were imported, therefore the tables are not described here.

2.1.8 Precipitation

Precipitation data are relevant for interpretation of atmospheric deposition measurements and for storm water runoff from impervious surfaces, Therefore the inventory contains a section to store metadata and data from precipitation gauges.

2.1.8.1 Table "precipitation"

This table provides time series of precipitation measurements.

Column name	Data type	Obligatory	Description
identifier_precipitation_gauge	text	x	identifier of the precipitation gauge, referring to table precipitation_gauge
observed_value	float8		Measured value, if the value was below LOQ or LOD the half of the respective value can be inserted.
start_interval_value	timestamptz	x	Value of the precipitation at the start of the referring intervall
interval_length	text	x	Length of the Intervall, Reference to the table "inerval_length"

2.1.8.2 Table "precipitation_gauge"

This table provides technical and spatial information about the gauge, that has been used for the precipitation measurements.

Column name	Data type	Obligatory	Description
identifier_precipitation_gauge	text	x	Identifier of the gauge. Must be unique in the table.
coordinates_longitude	float8	x	Value of the longitude coordinate of the coordinate reference system defined in column "coordinates_crs_epsg_code".
coordinates_latitude	float8	x	Value of the latitude coordinate of the coordinate reference system defined in "coordinates_crs_epsg_code".
coordinates_crs_epsg_code	int4	x	Coordinate reference system of the coordinates used in the two previous columns coded by the European Petroleum Survey Group

			(EPSG) code, e.g 4326 for WGS84 - CRS.
datasource_identifier	text	x	Identifier of the data source, referring to table “md_data_source” where details of all data sources are listed.

3 References...

Kosztra, B., Büttner, G., Hazeu, G., & Arnold, S. (10. 05 2019). *Updated CLC illustrated nomenclature guidelines.* Von https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/html/index.html abgerufen