TEXTE
163/2021

Final report

The database "Pharmaceuticals in the Environment"

Update for the period 2017-2020

by:

Stephanie Graumnitz, Dirk Jungmann
GWT-TUD GmbH, Dresden

publisher:
German Environment Agency

German Environment Agency

TEXTE 163/2021

Ressortforschungsplan of the Federal Ministry for the Environment,
Nature Conservation, Nuclear Safety and Consumer Protection

Project No. 146562

Report No. FB000627/ENG

Final report

The database "Pharmaceuticals in the Environment"

Update for the period 2017-2020

by

Stephanie Graumnitz, Dirk Jungmann
GWT-TUD GmbH, Dresden

On behalf of the German Environment Agency

Imprint

Publisher
Umweltbundesamt
Wörlitzer Platz 1
06844 Dessau-Roßlau
Tel: +49 340-2103-0
Fax: +49 340-2103-2285
buergerservice@uba.de
databasepharmaceuticals@uba.de
Internet: www.umweltbundesamt.de

/umweltbundesamt.de
/umweltbundesamt

Report performed by:
GWT-TUD GmbH
Freiberger Straße 33
01062 Dresden
Germany

Report completed in:
May 2021

Edited by:
Section IV2.2 - Pharmaceuticals
Arne Hein

Publication as pdf:
http://www.umweltbundesamt.de/publikationen

ISSN 1862-4804

Dessau-Roßlau, December 2021

Abstract: The database "Pharmaceuticals in the Environment" - Update for the period 2017-2020

Numerous publications worldwide dealing with the release of human and veterinary pharmaceutical from different emission sources into the environment. To overview this huge amount of information the German Environment Agency initiated a database project in 2014.

During two projects the database “Pharmaceuticals in the environment” (https://www.umweltbundesamt.de/en/database-pharmaceuticals-in-the-environment-0) was established (aus der Beek et al., 2016) and optimized (Dusi et al., 2019) and measured environmental concentrations (MEC) published until 2016 were inserted.

The aim of this project was the update of the database for MEC data of human and veterinary pharmaceuticals published between 2017 and 2020.

During the project 98,246 MECs of 543 publications, published between 2017 and 2020, were entered into the database. The complete database currently contains 276,895 MEC entries of 2062 publications from 89 countries. Additionally, 632 publications and 196 reviews published between 2017 and 2020 and dealing with pharmaceutical residues in the environment were added to an EndNote literature database.

Within the report, we provide evaluation on the analysed environmental matrices and the most prominent pharmaceutical residues detected in all UN regions.

Kurzbeschreibung: Die Datenbank „Arzneimittel in der Umwelt“ – Aktualisierung für den Zeitraum 2017-2020

Weltweit thematisieren zahlreiche Publikationen den Eintrag von Arzneimitteln aus verschiedenen Emissionsquellen in die Umwelt. Um diese enorme Datenmenge zu überschauen initiierte das Umweltbundesamt im Jahr 2014 ein Datenbankprojekt.

Im Rahmen von zwei Projekten wurde seitdem die öffentlich zugängliche Datenbank „Arzneimittel in der Umwelt“ (https://www.umweltbundesamt.de/en/database-pharmaceuticals-in-the-environment-0) etabliert (aus der Beek et al., 2016) und optimiert (Dusi et al., 2019) und weltweite gemessene Konzentrationen von Arzneimittelrückständen in der Umwelt, die bis 2016 veröffentlicht wurden, eingetragen.

Das Ziel dieses Projektes war die Aktualisierung der Datenbank mit Umweltkonzentrationen von Arzneimitteln, die im Zeitraum 2017 bis 2020 publiziert wurden.

Als Ergebnis einer Literaturrecherche wurden während des Projektes 98.246 Umweltkonzentrationen aus 543 Publikationen in die Datenbank eingetragen. Die aktualisierte Datenbank enthält damit derzeit 276.895 Einträge mit Umweltkonzentrationen aus 2.062 Publikationen, gemessen in 89 Ländern. Zusätzlich wurden 632 Publikationen und 196 Review-Artikel für den Zeitraum 2017-2020, die Konzentrationen von Arzneimittelrückständen in der Umwelt thematisieren, zu einer bestehenden EndNote-Literatur-Datenbank hinzugefügt.

Im Rahmen des Berichtes wird ein Überblick über die Umweltmatrices in denen Arzneimittelrückstände weltweit gemessen wurden und die am häufigsten gemessen Substanzen gegeben.

Table of content

List of figures

List of tables

List of abbreviations

LOD	Limit of detection
LOQ	Limit of quantification
MEC	Measured environmental concentration
POCIS	Polar organic chemical integrative sampler

Summary

Project aim

The aim of the project was to update the database "Pharmaceuticals in the environment" of the German Environment Agency for measured environmental concentrations of human and veterinary pharmaceuticals published between 2017 and 2020.

Content

The report describes the methodical approach for the literature survey and data collection, for the data entry into the existing database and the analysis of the updated database. Issues and remarks concerning the data entry are outlined. A general overview of the updated database is given in form of world maps, which represents the global distribution of publications and MECs in different environmental matrices.

Results

Literature survey

After application and evaluation of search criteria the literature research with the platform Web of Science resulted in 7,353 publications and 582 review articles published between 2017 and 2020 and dealing with pharmaceuticals in the environment. All these articles were transferred into the bibliographic database program Zotero and reviewed for relevant data. 1,184 publications and 196 review articles dealing with pharmaceuticals in the environment were found and transferred to the existing EndNote literature database.

Data entry into the existing database

The data entry was performed in two work steps. First the relevant data were introduced in a MYSQL project database via manual or semi-automated data input. In the second step the new data entries in the MYSQL project database were implemented into the original MS Access database.

During the project 98,246 MECs of 543 publications, published between 2017 and 2020, were entered into the database. The complete database currently contains 276,895 MEC entries of 2,062 publications from 89 countries. Additionally, 632 publications and 196 reviews published between 2017 and 2020, which are dealing with pharmaceutical residues in the environment were added to the EndNote literature database.

Overview of the worldwide distribution of pharmaceuticals in the environment

The statistical analysis of the updated database shows that most of the publications and reported MECs in the database are assigned to the countries China, Germany, Spain and United States of America. Most of the positive detections were reported for matrices allocated to the matrix-aggregation cluster liquid-immission, especially for the matrix "Surface Water - River/Stream". Second most frequent positive detections were reported for the matrix-aggregation cluster liquid-emission and within this most frequently for the matrix "WWTP effluent (treated)". Positive detections for solid emission or immission matrices were reported considerably more seldom.

Zusammenfassung

Projektziel

Ziel des Projektes war die Aufnahme veröffentlichter Umweltkonzentrationen von Arzneimittelrückständen in die bestehende Datenbank „Arzneimittel in der Umwelt" des Umweltbundesamtes für die Jahre 2017 bis 2020.

Inhalt

Der Bericht beschreibt die methodische Herangehensweise bei der Literaturrecherche und der Datensammlung, sowie beim Eintrag relevanter Daten in die bestehende Datenbank und der anschließenden Analyse der aktualisierten Datenbank. Dabei werden Problematiken und Optimierungsvorschläge im Hinblick auf die Dateneingabe diskutiert. Durch die kartografische Darstellung der globalen Verteilung von relevanten Publikationen und Umweltkonzentrationen von Arzneimitteln in verschiedenen Umweltmatrices wird ein Überblick über die aktualisierte Datenbank gegeben.

Results

Literaturrecherche

Nach der Anwendung und Evaluierung von Suchkriterien ergab die Literaturrecherche mit der Plattform Web of Science 7.353 Publikationen and 582 Review-Artikel, die zwischen 2017 und 2020 zur Thematik Arzneimittel und Umweltkonzentrationen in verschiedenen Umweltmatrices veröffentlicht wurden. Die gefundenen Artikel wurden in das Literaturverwaltungsprogramm Zotero importiert und hinsichtlich der Relevanz für die Datenbank überprüft. Die relevante Literatur wurde in die EndNote-Literaturdatenbank übertragen.

Übertragung relevanter Umweltkonzentrationen in die bestehende Datenbank

Die Übertragung relevanter Umweltkonzentrationen in die bestehende Datenbank erfolgte in zwei Schritten. Zuerst wurden alle Daten aus den relevanten Artikeln durch manuellen oder halb-automatischen Dateneintrag in eine MYSQL Projekt-Datenbank eingepflegt. Im zweiten Schritt wurde die MySQL Projekt-Datenbank in die originale MS Access-Datenbank übertragen.

Im Rahmen des Projektes wurden 98.264 Umweltkonzentrationen aus 543 Publikationen, die zwischen 2017 und 2020 veröffentlicht wurden, zur bestehenden Datenbank hinzugefügt. Die aktualisierte Datenbank beinhaltet derzeit 276,975 Einträge von Umweltkonzentrationen aus 2,062 Publikationen und 89 Ländern. Zusätzlich wurden 632 Publikationen und 196 Review-Artikel, die Konzentrationen von Arzneimittelrückständen in der Umwelt thematisieren und zwischen 2017 und 2020 publiziert wurden, zur EndNote-Literaturdatenbank hinzugefügt.

Überblick über das weltweite Vorkommen von Arzneimitteln in der Umwelt

Die statistische Analyse der aktualisierten Datenbank zeigt, dass die meisten Publikationen und Daten zu Umweltkonzentrationen für die Länder China, Deutschland, Spanien und die Vereinigten Staaten von Amerika gefunden wurden. Die meisten Funde (also Umweltkonzentrationen oberhalb der Nachweisgrenze) konnten in Umweltmatrices verzeichnet werden, die der Gruppe „flüssig Immission" zugeordnet werden, allen voran der Matrix „Oberflächenwasser-Fließgewässer". Am zweithäufigsten wurden Substanzen in Matrices, die der Gruppe „flüssig Emission" zugeordnet werden, detektiert. In dieser Gruppe konnten die häufigsten Funde in der Matrix „Kläranlageablauf" verzeichnet werden. Detektionen von Substanzen in den Matrices der Gruppen „fest Emission" und „fest Immission" wurden deutlich seltener publiziert.

1 Introduction

Numerous publications worldwide dealing with the release of human and veterinary pharmaceutical from different emission sources into the environment. The high number of publications concerning this issue indicate the importance of this topic, but the steadily increasing number of publications made it difficult to oversee the global occurrence of pharmaceutical residues, main emission sources, effects on the environment and possible problem-solving approaches. That in this respect a global consideration is necessary became at least apparent when the global policy framework “SAICM-Strategic Approach to International Chemicals Management” was decided.

In this context, the German Environment Agency initiate the project "Pharmaceuticals in the environment - occurrence, effects and options for action" in 2014. In this project worldwide measured environmental concentrations (MECs) of human and veterinary pharmaceutical and their transformation products were compiled in a database (aus der Beek et al., 2016). The aim of the subsequent project “The database `Pharmaceuticals in the Environment´ – Update and new analysis” was the update and optimization of the existing database concerning worldwide MECs published between 2010 and 2016 (Dusi et al., 2019).

The database is growing in appeal and is used and cited in different publications (Olarinmoye et al., 2015, Emara et al., 2018, Booth et al., 2020, Jameel et al., 2020). Since Spring 2021 the database is also publicly available over the Information Platform for Chemical Monitoring (IPCHEM, https://ec.europa.eu/newsroom/ipchem/items/701651). Hence, a regularly update of the database is useful and necessary to offer an updated overview to the users.

The current project aims at the update of the existing database for worldwide MECs published between 2017 and 2020. Therefore, a literature research and review were performed. The MEC values found during this literature survey were inserted into the existing database and the existing literature database was updated. Furthermore, an analysis of the updated database was performed, and selected results were visualised in world maps.

2 Project aim

The aim of the project was the update of the database "Pharmaceuticals in the environment" of the German Environment Agency. To attain this aim following work stages were realized:

- Application and evaluation of search and exclusion criteria for the literature research of publications using the platform Web of Science
- Review of the publications found during the literature research
- Implementation of worldwide MECs of human and veterinary pharmaceuticals from studies published between 2017 and 2020 into the existing MS Access database of the German Environment Agency
- Implementation of relevant publications in an EndNote literature database
- Visualization of selected data as world maps

3 Literature survey and Data collection

3.1 Protocol of literature research and defining search parameters.

For the literature research the platform Web of Science was used. Most databases listed in Web of Science were selected for the research, including Web of Science Core Collection, BIOSIS Citation Index, BIOSIS Previews, Current Contents Connect, Data Citation Index, Korean Journal Database (KCI), MEDLINE®, Russian SciELO Citation Index and Zoological Record. Only Derwent Innovations Index and the document type "Patent" were excluded from search as they focus on patent searching.

The literature research was performed for literature published between 2017 and 2020, since earlier published date have already been analysed by aus der Beek et al. (2016) and Dusi et al. (2019). The same definition of "pharmaceutical" was used as described in aus der Beek et al. (2016) and Dusi et al. (2019), where pharmaceuticals are defined as substances that are mainly used for therapeutic purpose. For this reason, substances used in personal care products (e.g., DEET in insect repellents or octocrylene in sunscreens), natural substances (e.g., thymol) as well as legal and illegal drugs, only used for recreational purpose (e.g., caffeine, nicotine, cocaine, MDA and MDMA) have not been considered for the database.

A specific query was used to limit the results of the literature survey to the above-mentioned search parameters. Thereby it was refrained from the use of exclusion criteria, like the exclusion of research areas, to avoid the unforeseen exclusion of relevant publications. The final query was performed using following search items with the operators provided by Web of Science:

*TS= (*pharmaceutical* AND (surface NEAR/1 water OR river OR stream OR running NEAR/1 water OR lake OR reservoir OR drinking NEAR/1 water OR tap NEAR/1 water OR soil OR sediment OR manure OR ground NEAR/1 water OR waste NEAR/1 water OR sewage) AND (occurrence OR concentration OR distribution OR monitoring OR presence)) AND PY= (2017-2020)*

This query resulted in 7935 publications including 582 review articles (Table 1).

Table 1: Result of the final literature query.

	2017	2018	2019	2020
Publications (review articles excluded)	1587	1780	2033	1953
Review articles	112	125	151	194
Total	**1699**	**1905**	**2184**	**2147**

3.2 Article review and bibliographic database

All 7,935 publications were transferred to the bibliographic database program Zotero (Zotero ©, Corporation for Digital Scholarship and Roy Rosenzweig Center for History and New Media, Virginia, USA).

The titles and abstracts were scanned for MEC-data and sorted in collections and subcollections listed in Table 2. Duplicated literature entries (about 100 publications) were deleted.

Table 2: Result of the article review

Literature database collection	Description of collection	Number of publications
MEC_inserted in database	Publications containing MEC-data and with entries in the database	543 (+5 associated articles)
manual data-import	Publications, where MEC-data were imported via manual data-import	333
semi-automated data-import	Publications, where MEC-data were imported via semi-automated data-import	210
MEC_not inserted in database	Publications containing MEC-data, but without database entry	145
MEC_data not applicable	Publications containing MEC-data in a non-applicable form for the database (e.g. charts, foreign languages, POCIS data)	154
MEC_no access	Publications without access, most probably containing MEC-data	189
MEC_no access to supplementary	Publications were access to the supplementary is needed but not given	144
unsure_no access	Publications without access, most probably not containing MEC-data	222
not relevant	Publications without relevant MEC-data	5811
not relevant, but related	Publications which are not relevant for the database but thematically related	47
review article		582
review article, relevant	Review articles dealing with pharmaceutical residues in the environment	195
review article, not relevant	Not relevant review article	387

During the review process, access for some publications was limited. Therefore, the access to these publications, which would most likely contain MEC-data, were kindly offered by employees of the German Environment Agency, Dessau-Roßlau.

3.3 Evaluation of the literature survey

For the evaluation of the literature survey the latest review articles were used. Five review articles published in 2020 (Aryal et al., 2020, Nassour et al., 2020, Ngqwala & Muchesa, 2020, Petrie & Camacho-Munoz, 2020, Valdez-Carrillo et al., 2020) were randomly chosen and the reference lists of these articles were screened for relevant article, thus articles published between 2017 and 2020 dealing with pharmaceutical in the environment. Relevant articles found in the reference lists of the five review articles were compared with the result of the literature survey. The result of this comparison is shown in Table 3. Around 90 % of the relevant articles cited in the five chosen review articles were found during our literature research with the platform Web of Science. One reason for the mismatch between cited literature in review articles and the literature research in this project is the used search algorithm. One example is the relevant article Bedoya-Rios et al. (2018) cited in Valdez-Carrillo et al. (2020), which was not found during our literature research. The reason is, that the title, abstract and key words of Bedoya-Rios et al. (2018) indeed contain the words occurrence and water, but not in combination with the word pharmaceutical, which is part of the search algorithm. This issue could be solved if the search algorithm is expanded, e.g. by specific names of substance classes. But in doing so the share of relevant publications on the total number of publications found during literature search should keep in mind to minimize the time-consuming sorting of relevant and non-relevant literature.

Table 3: Comparison of the literature cited in the latest review articles and the result of the literature research

	No. of relevant articles cited in the review articles	No. of relevant articles recovered during literature research	% of relevant articles recovered during literature research
Aryal et al., 2020	13	11	85
Nassour et al., 2020	3	3	100
Ngqwala & Muchesa, 2020	16	16	100
Petrie & Camacho-Munoz, 2020	15	14	93
Valdez-Carrillo et al., 2020	30	26	87

4 Creating the database and analysing data

4.1 Technical background of the database

As already described in Dusi et al. (2019) the original database "Pharmaceuticals in the Environment" is provided as a Microsoft-Access database. Due to reasons of practicability Dusi et al. (2019) decided "to reimplement a normalized version of the original DATABASE utilizing state of the art open-source web technology by a LAMP-server (Linux – Ubuntu 16.04 LTS, Apache 2.4.18, MySQL 5.7.23, PHP 7.0.32) hosted within the campus net of the TU Dresden. User - DATABASE interactions were provided utilizing a specific PHP based web frontend and the MySQL Workbench (v. 6.3)". We used the same method and entered the MEC-data into a MySQL project DATABASE. At the end of the project the data from the MySQL project database were transferred back to the Microsoft-Access database. As described in Dusi et al. (2019) the data transfer between the MySQL database and the Microsoft-Access database was performed using R (version 4.0.3, R Core Team, 2020) and the packages RMySQL (Ooms et al., 2020) and RODATABASEC (Ripley and Lapsley, 2020).

4.2 Import of huge data sets

Especially for publications with huge data sets Dusi et al. (2019) developed a semi-automatic approach of data import into the MySQL project database. As this approach decreases the probability of mistakes and is considerably time saving it was also used in this project. According to Dusi et al. (2019) the semi-automatic data import consists of the following work steps:

1. **Harmonization:** Transfer of raw data into Excel and transformation into a generalised data format
2. **Data Completion:** Assignment of required and available data as well provision of further required information e.g. from the text
3. **Database Preparation:** Creation of new entries in sub tables if required.
4. **Data Import:** Test, final corrections and data transfer

In this project we used the same general but flexible framework consisting of two worksheets in an Excel-file. In the first worksheet (meta table) all meta information required to create an entry within the MySQL project database were implemented. The second worksheet (value table) contains a harmonized copy of the original database table, where the sampling information, MEC values, LOD values and possibly statistic-information for each single MEC value were implemented.

For the data import we used a specific import script based on R (version 4.0.3, R Core Team, 2020), as well. The import script was adjusted to enable the import of MEC values with different statistic information (e.g. "single value" and "smaller than").

For further information concerning the semi-automated data import see Dusi et al. (2019).

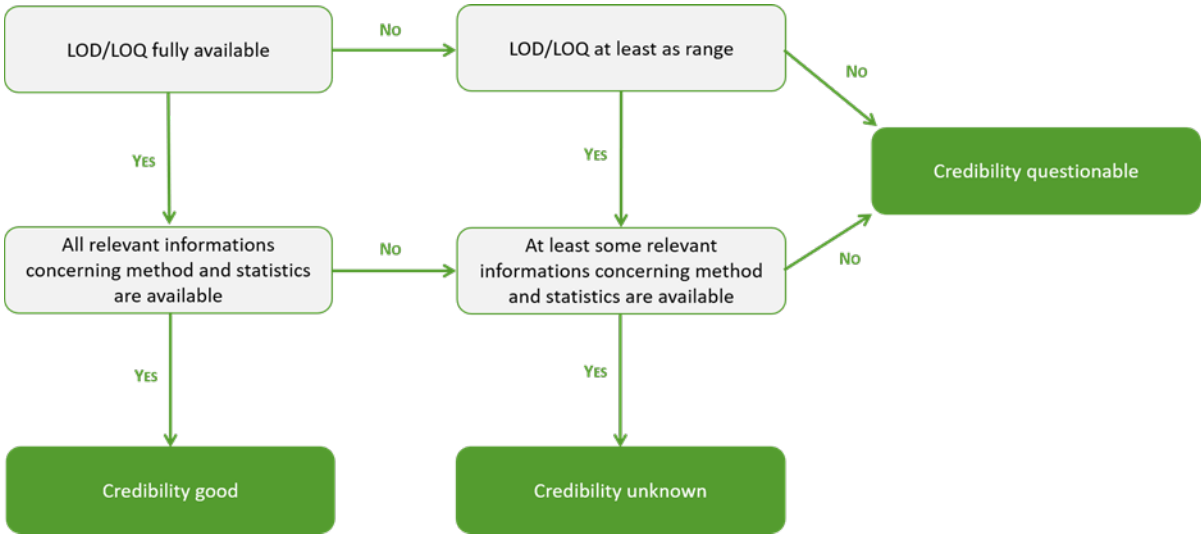
4.3 Issues during data entry and further remarks

During the process of data entry into the MySQL project database and the subsequent data transfer to the MS Access database we became aware of some issues, partially already discussed in Dusi et al. (2019). In the following these issues are described and possible solutions are proposed.

4.3.1 Remarks concerning the table "literature"

- The attribute *credibility of reference* with its categories "good", "questionable" and "unknown" is quite subjective. Therefore, we tried to find an intern standard decision guidance, shown in Figure 1. It should be considered that this is not an alternative or second peer review but should be seen as categorization of the suitability of the respective publication for the database. Furthermore, it should draw user's attention concerning possible disagreements or missing attributes in the publications. The intern standard guidance is a first draft and should be seen as basis for further discussion concerning this issue.

Figure 1: Intern decision guidance for the categorization of publications concerning their credibility.



Source: own illustration, GWT-TUD GmbH

- For more consistency fixed categories should be allocated to the attribute research type. We suggest the categories listed in Table 4.

Table 4: Possible categories for the database attribute research type

Category	Description
Cooperation Research	Cooperation of institutions related to different categories
Environmental Institute Research	Research of institutes with focus on environmental research
Federal/State Agency Research	Research of public institutions/agencies
Industrial Research	Research of companies
University Research	Research of universities or further institutions of higher education
Scientific Institute Research	Research of institutes, independent of the scientific orientation

4.3.2 Remarks concerning the attributes "Detection" and "Limit of Detection"

As already mentioned in Dusi et al. (2019) there are still inconsistencies in the database concerning the attributes "Detection" and "Limit of Detection" which remained unresolved as they were not addressed during this project.

The attribute "detection" with its categories "below detection" (MEC<LOD) and "positive detection" (MEC>LOD) is insufficient, due to the plenty of variations how limits of chemical analysis are reported in publications. An excerpt of this plenty of variation and the consequential database entries are listed in Table 5.

Table 5: List of wordings for detection limits reported in investigated publications and their correspondence in the database

	Publication		Database		
	Data Table	**Limit of detection**	**Substance Detection**	**MEC value**	**Limit of Detection**
Manuell data import	n.d.	yes	below detection	0	*value LOD*
	<LOD	yes	below detection	0	*value LOD*
	<LOQ	yes	positive detection	<-9999	*value LOD*
	< *value LOD*	yes	below detection	0	*value LOD*
	< *value LOQ*	yes	positive detection	< *value LOQ*	*value LOD*
	n.d.	no	below detection	0	-9999
	<LOD	no	below detection	0	-9999
	<LOQ	no	positive detection	<-9999	-9999
	< *value LOD*	no	below detection	0	-9999
	< *value LOQ*	no	positive detection	< *value LOQ*	-9999
Semi-automated data import	n.d.	yes	below detection	0	*value LOD*
	<LOD	yes	below detection	0	*value LOD*
	<LOQ	yes	positive detection	<-9999	*value LOD*
	< *value LOD*	yes	below detection	0	*value LOD*
	< *value LOQ*	yes	positive detection	< *value LOQ*	*value LOD*
	n.d.	no	below detection	0	-9999
	<LOD	no	below detection	0	-9999
	<LOQ	no	below detection	<-9999	-9999
	< *value LOD*	no	below detection	0	-9999
	< *value LOQ*	no	positive detection	< *value LOQ*	-9999

- A related issue is the difference of the interpretation of the attribute "detection" between the manual data import and the semi-automated data import. This problem has been clearly outlined in Dusi et al. (2019) and becomes apparent from Table 5 (see bold type). During the manual data entry missing LOD-values do not necessarily lead to the result "below detection". If the authors distinguish in MEC-reporting between <LOD and <LOQ, the result <LOD is rated as "below detection", and the result <LOQ as "positive detection" with the MEC value <-9999. During the semi-automated data import there is stricter definition, where the attribute "Detection" is calculated by logical comparison of standardized MEC and standardized LOD. Therefore, the MEC-reporting <LOQ results in the MEC entry <-9999 and "below detection". Only if a value for LOQ and LOD is reported a <LOQ results in "positive detection". Therefore, results filtered by the attribute detection should interpreted with care.

- Beside the variation reported in Table 5 there are lot of additional terms (e.g. *method detection limit*, *method quantification limit*, *method reporting limit*, *instrument detection limit*, *lowest calibration level*) which are partially synonyms of the terms LOD and LOQ but partially deviating terms underlying complete different statistical or analytical methods.
- We recommend extending the database for the attribute "analytical limit" at least with the categories "limit of detection", "limit of quantification" and "unclear". Consequential and already recommended in Dusi et al. (2019) the category "unclear detection" should be added to the attribute "Detection".
- It occurs that authors decide, to present only MEC values of substances with at least one "positive detection" in the data tables of their publications. This is worth mentioning as it leads to a shift in the relationship between measured and detected substances.

4.3.3 Remarks concerning the attribute "Statistic"

- The currently implemented categories of statistics are partially inappropriate for new sampling methods. This applies particularly to the method of passive sampling. Therefore, we added the statistical category "time-weighted-average concentration" to the attribute "Statistics". This works of course only if results of passive sampling are reported as time-weighted-averages with a unit, which could convert to the standardized unit µg/L. Results reported as concentration/sampler cannot be inserted into the database.
- Users of the database should consider that there is no possibility to identify whether median or mean values reported in publications are calculated from technical replicates (1 sample was analysed in duplicate or triplicate) or based on the aggregation of single MEC values of several samples without consideration of the attributes "number of samples" and if applicable "Description".

4.3.4 Further remarks

For further optimization of the Access-database we recommend:

- The implementation of a column for comments. This is for example useful to statement why the credibility of a publication is not rated with "good". A separate literature table within the Access-Database with such a comment column would be the best solution.
- The implementation of a separate column for the publication year. This would make it much easier to filter for a special timespan.

4.4 Data analysis and map creation

Data analysis and visualization was performed using the script language R (version 4.0.3, R Core Team, 2020) and the packages RMySQL (Ooms et al., 2020) and RODATABASEC (Ripley and Lapsley, 2020) for database communication.

For geographical data visualization, the GIS-world map TM_WORLD_BORDERS-0.39 was used. As interface between GIS and the MEC database the ISO2-Codes of countries Codes (DIN EN ISO 3166-1) were utilized. Maps were created using the R packages maptools (Bivand and Lewin-Koh, 2021), rgeos (Bivand and Rundel, 2020) and rgdal (Bivand, et al., 2021) as well as RColorBrewer (Neuwirth, 2014) for the creation of colour gradients. The map creation was performed according to Dusi et al. (2019). For the maps 1 and 2 (Figure 2 and Figure 3) the data were aggregated for the country name respectively the ISO2 code. For the maps 3-6 (Figure 4 to Figure 7) the data entries were assigned to the categories liquid emission, liquid immission, solid emission and solid immission or excluded in respect of the origin (emission or immission) and aggregate state (liquid or solid) of their matrices (see Table 6). Subsequently, the data entries were aggregated for the country name respectively the ISO2 code and filtered for data entries with “positive detection”.

5 General analysis of the database

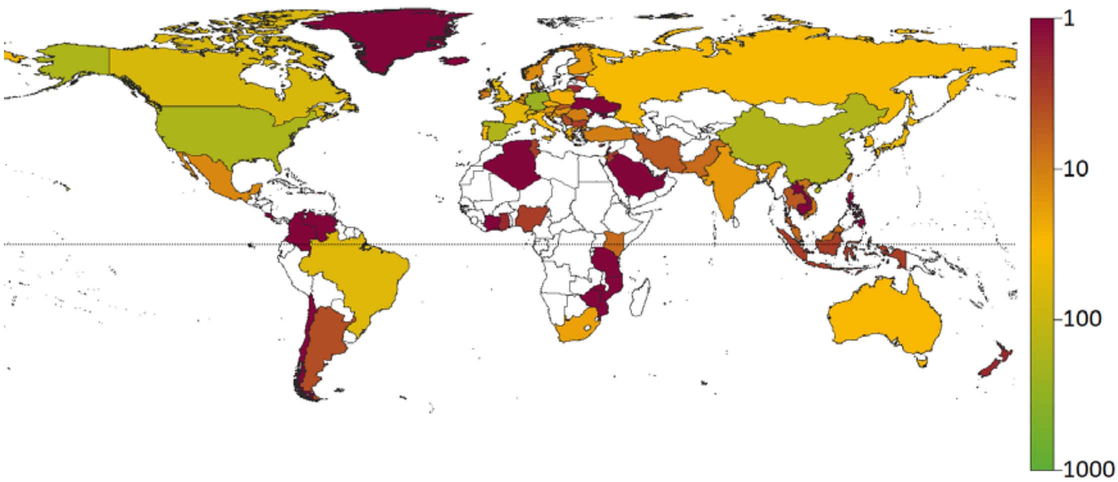
5.1 General overview of the database

In total 98,246 MEC data of 543 publications, published between 2017 and 2020, were entered into the database during this project. The complete database contains currently 276,895 MEC entries of 2,062 publications. Additionally, 632 publications and 196 reviews reporting MEC-data were added to the literature database.

5.2 Global distribution of MEC-reporting literature

The current state of the database "pharmaceuticals in the environment" contains MEC-data reported from 89 countries. For the timespan 2017-2020 ten further countries reporting MEC-data, namely Bangladesh, Cameroon, Egypt, Latvia, Lesotho, Moldova, Republic of Iraq, Republic of Zambia, Sri Lanka and Uruguay, were added to the database. The worldwide distribution of countries reporting MEC-data is presented in Figure 2. Green colours indicate countries with high numbers of publications reporting MEC-data, red colours indicate low numbers of publications reporting MEC-data.

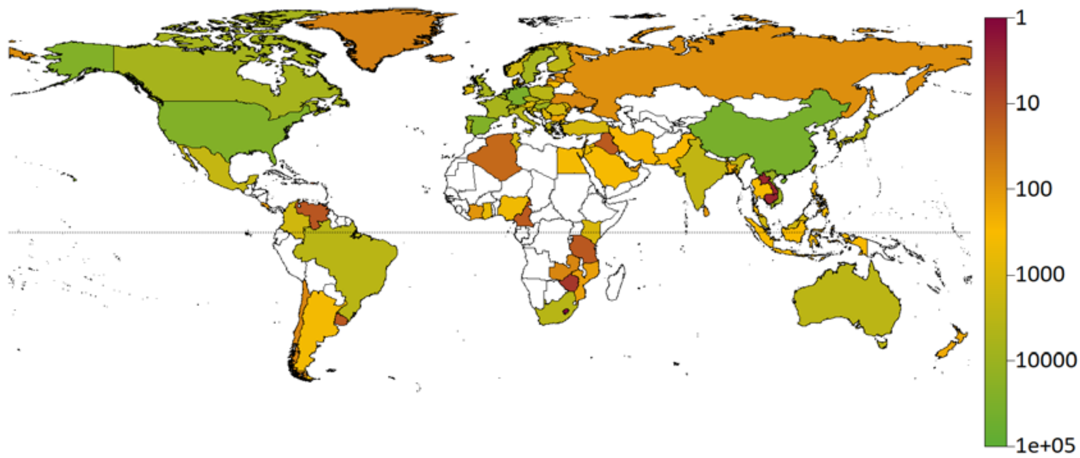
Figure 2: Global distribution of publications reporting MEC-data published until 2020



Source: own illustration, GWT-TUD GmbH

The most of the 2062 publications in the database report data from Germany (295 publications), China (274 publications), United States of America (226 publications) and Spain (199 publications). The global distribution of the published MEC-data is shown in Figure 3. Most of the 276,895 MEC entries in the database, are reported for China (37,708 MEC entries), Germany (34,001MEC entries), Spain (26,988 MEC entries) and USA (25,647 MEC entries), as well. The latter measurement has to be interpreted with care, as the MEC entries are values of different aggregation levels (single value, median or maximum).

Figure 3: Global distribution of MEC values published until 2020



Source: own illustration, GWT-TUD GmbH

5.3 Distribution of pharmaceuticals in different environmental matrices

To get an overview concerning the distribution of pharmaceuticals in different environmental matrices, the different environmental matrices reported in the publications were classified in respect to the aggregate status (liquid or solid) and if they are sources of pharmaceutical release (emission) or influenced by pharmaceutical release (immission). This classification follows the classification described in Dusi et al. (2019). During the data import of the years 2017-2020 further emission sources were added to the database (see Table 6, bold type).

- Compost
- Sediment lagoon
- Sewage urban (treated): Sewage, which was not treated in a wastewater treatment plant, but passed a kind of pre-treatment bevor release into the environment
- Raw water - Drinking water treatment plant: Raw water of a drinking water treatment plant, when the source of the raw water (e.g. ground water, spring water) was not specified
- Spring water: water from a spring
- Sewage (untreated) – unspecific: untreated sewage which could not be assigned to a specific source (urban, industrial, hospital)
- Surface water – Pond: ambiguous definition, natural ponds and/or fishponds, excluded from statistics, because an assignment to emission or immission is not possible

While in Table 6 all values of the database are included, matrices with inadequate description (e.g. matrices “unknown” or Surface water – unspecific), matrices which could not assigned to one of the summarized categories (e.g. “Leachate”, “Surface water – Pond”) as well as inflow and internal sewage treatment steps of WWTPs or values without unit were excluded from further analysis. The Matrix “Soil water” also includes samples from “Infiltration water” of lake/river sediments.

The most MEC values and the most MEC values with positive detection in the database (see Table 6) were reported for the Matrices "Surface water - River/Stream" (36,162 positive detection entries), "WWTP effluent (treated)" (26,567 positive detection entries) and "WWTP inflow (untreated)" (16,089 positive detection entries).

Most positively detected MECs in the database have been reported for matrices clustered within liquid-immission (40.9 %) followed by the matrices clustered in liquid-emission (27.0 %). Positively detected MECs measured in matrices which are clustering in solid-emission and solid-immission count only for 3.9 % and 7.9 % of total number of positively detected MEC entries. Twenty percent of the positively detected MEC entries were not assigned to one of the four matrix-clusters, thus excluded from analysis.

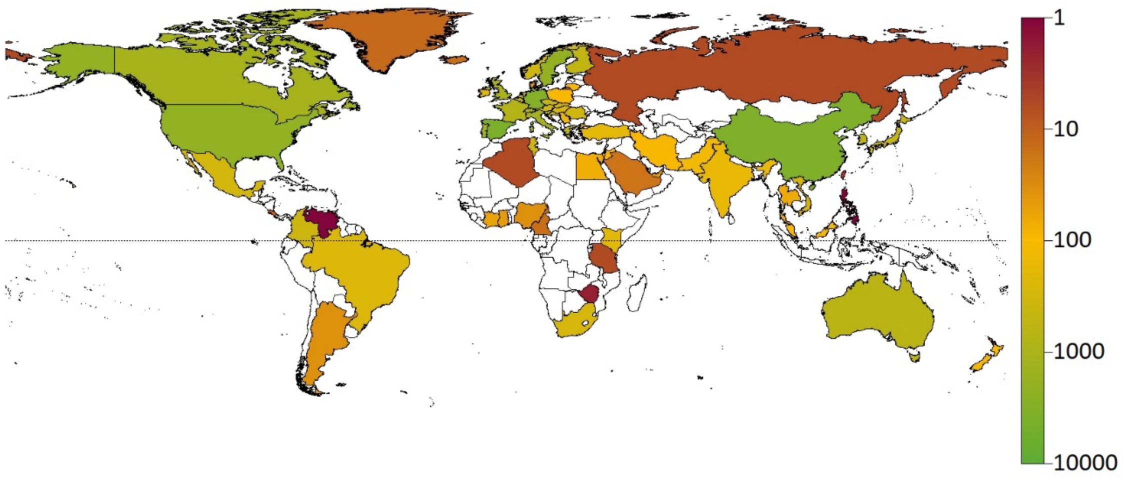
Table 6: Different matrices of the database and their classification in aggregate status and emission and immission classification as well as the number of MECs with positive detection (>LOD) and all MECs reported.

Aggregate Status	Immission/Emission	Environmental Matrix	Number of positive MEC entries	Number of MEC entries
Liquid	Human emission	Reclaimed water	21	53
		Sewage hospital (treated)	516	798
		Sewage hospital (untreated)	2,634	3,866
		Sewage industrial (untreated)	412	873
		Sewage urban (untreated)	1,450	2,614
		Sewage urban (treated)	73	180
		WWTP effluent (treated)	26,567	51,860
	Veterinary emission	Manure – liquid	476	1003
		Sewage livestock (treated)	45	88
		Sewage livestock (untreated)	324	585
		Surface water – Aquaculture	279	1,123
	Immission	Drinking water	1,952	7,557
		Groundwater	4,368	19,636
		Raw Water - Drinking Water Treatment Plant	386	526
		Reservoir drainage	10	27
		Riverbank filtration	207	1,011
		Soil water	315	1,045
		Spring water	10	68
		Surface water - Estuary	1,574	3,263
		Surface water - Lake	2,892	6,431

Aggregate Status	Immission/Emission	Environmental Matrix	Number of positive MEC entries	Number of MEC entries
		Surface water - River/Stream	36,162	104,391
		Tap water	634	1,673
		Well water (untreated)	1,093	4,845
	excluded	Dissolved activated sludge	6	54
		Leachate	417	614
		Rain	45	72
		Sea ice	0	10
		Sewage (untreated) - unspecific	83	210
		Surface water - Sea or Ocean	2,081	4,264
		Surface water - Pond	77	123
		Surface water - unspecific	4,226	6,514
		WWTP desinfection effluent	94	117
		WWTP inflow (untreated)	16,089	26,039
		WWTP primary effluent	88	216
		WWTP secondary effluent	357	536
		Unknown	171	358
Solid	Human emission	Compost	51	51
		Sewage sludge	51	68
		Suspended particulate matter - Sewage	96	280
		Suspended particulate matter (WWTP-Effluent)	17	17
		WWTP biosolid	316	460
		WWTP dehydrated sludge	25	86
		WWTP digested sludge	132	189
		WWTP sludge	2,919	4,343
	Veterinary emission	Manure - dung	992	1,635
		Sediment - Aquaculture	120	235
	Immission	Sediment - Estuary	210	537
		Sediment - Lagoon	20	22
		Sediment - Lake	218	981

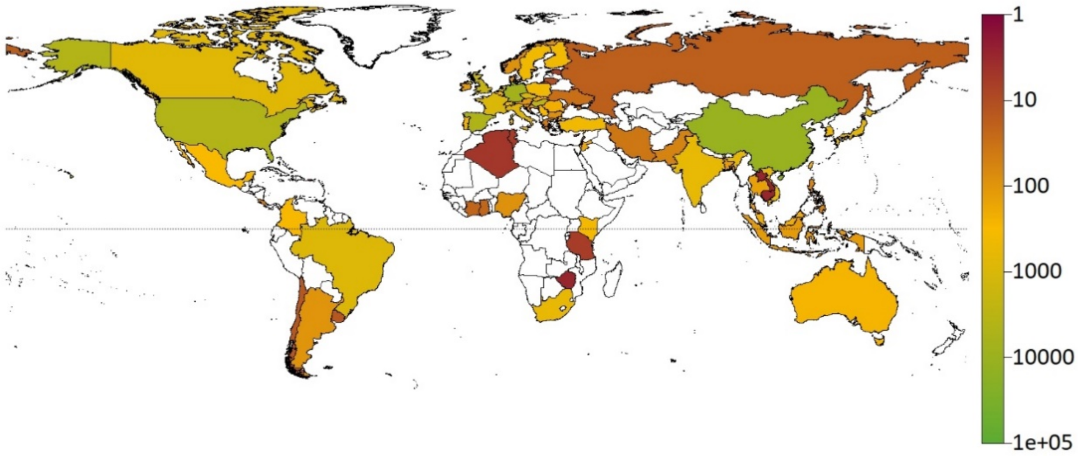
Aggregate Status	Immission/Emission	Environmental Matrix	Number of positive MEC entries	Number of MEC entries
		Sediment - River/Stream	1,996	4,964
		Soil	1,556	2,678
		Suspended particulate matter - Estuary	65	180
		Suspended particulate matter - River/Stream	5,490	6,025
	excluded	Dust	12	18
		Sediment - Sea or Ocean	132	525
		Sediment - unspecific	370	543
		Suspended particulate matter - Sea or Ocean	5	12
		Suspended particulate matter - unspecific	107	111
		Suspended particular matter (WWTP-Inflow)	84	122
		WWTP primary sludge	44	60
		WWTP secondary sludge	90	110
Sum			**121,252**	**276,895**

Figure 4: Global distribution of MEC entries with positive detection in liquid emission matrices



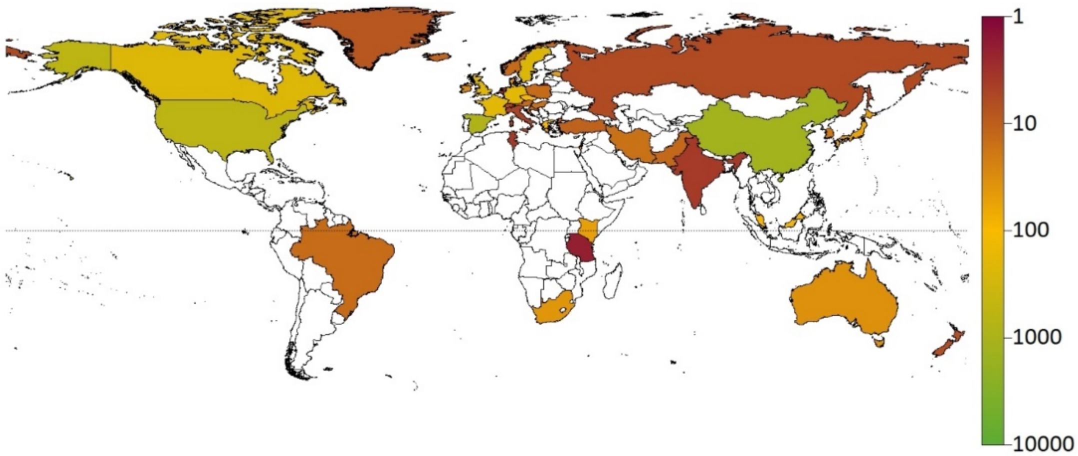
Source: own illustration, GWT-TUD GmbH

Figure 5: Global distribution of MEC entries with positive detection in liquid immission matrices



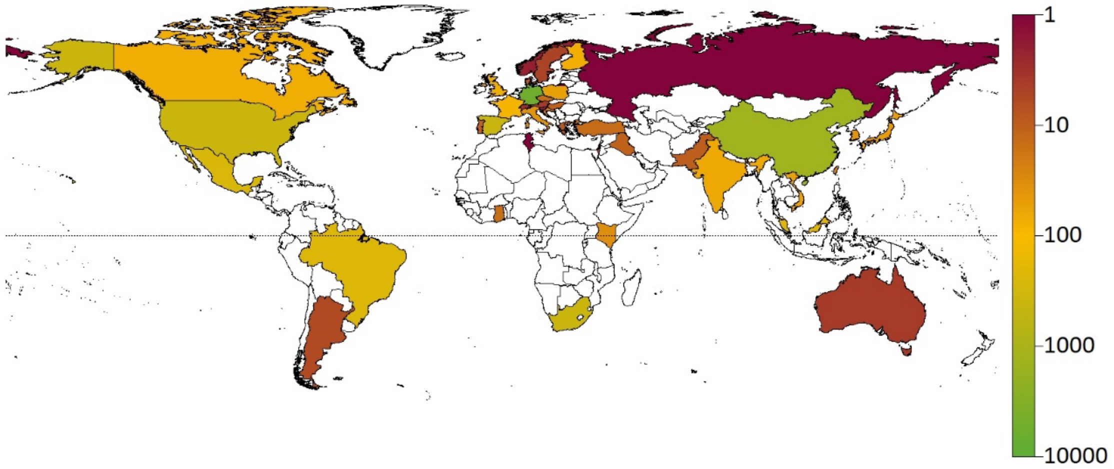
Source: own illustration, GWT-TUD GmbH

Figure 6: Global distribution of MEC entries with positive detection in solid emission matrices



Source: own illustration, GWT-TUD GmbH

Figure 7: Global distribution of MEC entries with positive detection in solid immission matrices



Source: own illustration, GWT-TUD GmbH

The global distribution of MEC entries with positive detection for the four environmental matrix clusters are shown in Figure 4-Figure 7. The most positively detected MEC entries for liquid emission matrices were reported for Spain, China, Germany and USA (Figure 4). For liquid immission matrices the most positively detected MEC entries were reported for China, Germany, Spain and USA. The most positively detected MEC entries in solid emission matrices (Figure 6) were reported for China, Spain, USA and Sweden. In solid immission matrices most positively detected MEC entries were reported for Germany, China, Spain and USA (Figure 7).

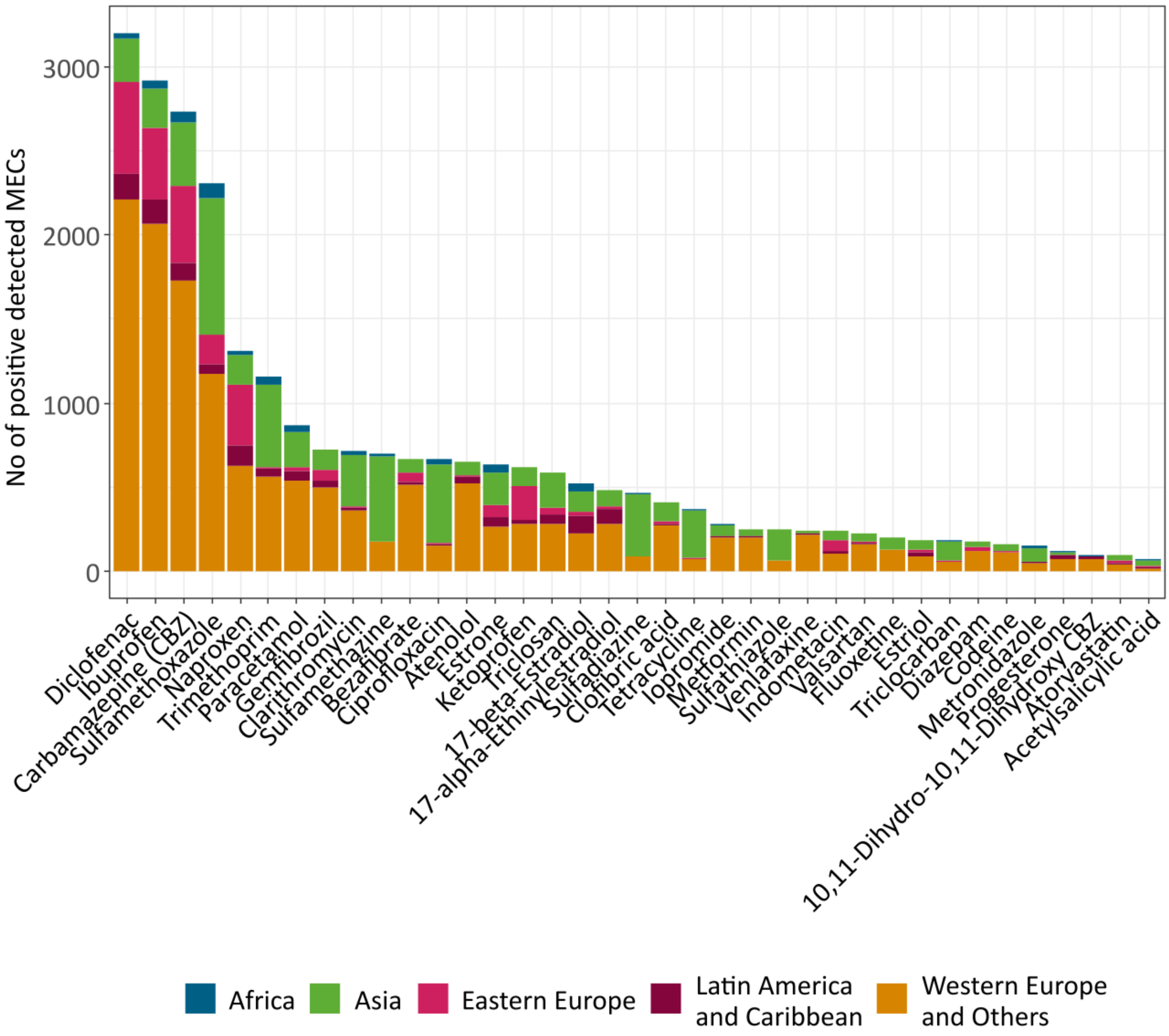
Table 7: Overview of the relevant numbers on a global scale, in the EU and Germany

Number of	Global	European Union	Germany
Publications	2,062	940	295
MECs	276,895	151,633	34,001
Positively detected MECs	121,252	61,618	17,693
Liquid emission	32,561	19,018	3,011
Liquid immission	49,118	24,426	7,540
Solid emission	4,629	2,061	180
Solid immission	9,522	6,248	5,414
Positively detected substances	992	749	414
Liquid emission	771	591	339
Liquid immission	703	483	198
Solid emission	337	250	39
Solid immission	295	227	120

To get an overview on the most prominent pharmaceuticals in the liquid matrices, we analysed the occurrence of all pharmaceuticals focusing on their occurrence in the five UN-regions. The matrices 'Groundwater', 'Spring water', 'Drinking Water', 'Surface Water - Estuary', 'Surface Water - Lake', 'Surface Water - River/Stream', 'Raw Water - Drinking Water Treatment Plant', 'Tap Water' and 'Well Water (untreated)' were considered for the analysis.

Thereby, we found 37 substances and metabolites of substances occurring in all five UN-regions. In comparison to Dusi et al. (2019) the 18 substances Atenolol, Atorvastatin, Bezafibrate Clarithromycin, Codeine, Diazepam, Fluoxetine, Gemfibrozil, Iopromide, Metformin, Metronidazole, Progesterone, Sulfadiazine, Sulfathiazole, Tetracycline, Valsartan, Venlafaxine and 10,11-Dihydro-10,11-Dihydroxy Carbamazepine, listed in Figure 8 and Table 8 in decreasing order of the total amount of positive detections, have now been reported with positive detections for all five UN-regions.

Figure 8: Number of positively detected MECs in surface water, groundwater or drinking water for substances occurring in all five UN-regions.



Source: own illustration, GWT-TUD GmbH

6 Conclusion and outlook

The aim of the project was the update of the database “Pharmaceuticals in the environment” of the German Environment Agency for measured environmental concentrations of human and veterinary pharmaceuticals published between 2017 and 2020.

The applicated search algorithm for the literature research with Web of Science resulted in a relative high amount of 7,353 publications and 582 review articles which had to be reviewed for relevant data. 1,184 publications and 196 review articles dealing with pharmaceutical residues in the environment were found and transferred to the existing EndNote literature database. During the project 98,246 MECs of 543 publications, published between 2017 and 2020, were entered into the database. The complete database currently contains 276,895 MEC entries of 2062 publications from 89 countries.

During this project we used the MySQL project database and the PHP based web frontend, as well as the excel data sheet and the R script for the data import of huge data sets developed by Dusi et al. (2019). This technical realization is particularly useful, especially if several people work on the database. Nevertheless, it became clear, that a manual how to enter data into the DATABASE is necessary. There are a lot of attributes in the database with no or vague definitions open to interpretation. Therefore, we share the recommendation of Dusi et al. (2019) for the development of a guideline for a standardized input of data into the database. The remarks concerning different issues of data input outlined in Dusi et al. (2019) and this report may serve as basis for such a guideline.

The updated database will also be integrated into the Information Platform for Chemical Monitoring - IPCHEM (link: https://ipchem.jrc.ec.europa.eu/).

List of references

Monographs:

aus der Beek, T., Weber, F.-A., Bergmann, A., Grüttner, G., Carius, A. (2016): Pharmaceuticals in the environment: Global occurrence and potential cooperative action under the Strategic Approach to International Chemicals Management (SAICM). UBA Texte 67, Umweltbundesamt Dessau, Deutschland https://www.umweltbundesamt.de/sites/default/files/medien/1968/publikationen/iww_abschlussbericht_saicm_arzneimittel_final.pdf

Dusi, E., Rybicki, M., Jungmann, D. (2019): The database "Pharmaceuticals in the Environment" - Update and new analysis. UBA Texte 67. Umweltbundesamt Dessau, Deutschland. https://www.umweltbundesamt.de/sites/default/files/medien/1410/publikationen/2019-06-24_texte_67-2019_database_pharmaceuticals-environment_0.pdf

Articles taken from journals:

Aryal, N., Wood, J., Rijal, I., Deng, D., Jha, M. K., & Ofori-Boadu, A. (2020). Fate of environmental pollutants: A review. Water Environment Research. https://doi.org/10.1002/wer.1404

Bedoya-Rios, D. F., Lara-Borrero, J. A., Duque-Pardo, V., Madera-Parra, C. A., Jimenez, E. M., & Toro, A. F. (2018). Study of the occurrence and ecosystem danger of selected endocrine disruptors in the urban water cycle of the city of Bogota, Colombia. Journal of Environmental Science and Health Part A-Toxic/Hazardous Substances & Environmental Engineering, 53(4), 317–325. https://doi.org/10.1080/10934529.2017.1401372

Booth, A., Aga, D. S., & Wester, A. L. (2020). Retrospective analysis of the global antibiotic residues that exceed the predicted no effect concentration for antimicrobial resistance in various environmental matrices. Environment International, 141, 105796.

Emara, Y., Lehmann, A., Siegert, M. W., & Finkbeiner, M. (2019). Modeling pharmaceutical emissions and their toxicity-related effects in life cycle assessment (LCA): A review. Integrated environmental assessment and management, 15(1), 6-18.

Jameel, Y., Valle, D., & Kay, P. (2020). Spatial variation in the detection rates of frequently studied pharmaceuticals in Asian, European and North American rivers. Science of the Total Environment, 724, 137947. https://doi.org/10.1016/j.scitotenv.2020.137947

Nassour, C., Barton, S. J., Nabhani-Gebara, S., Saab, Y., & Barker, J. (2020). Occurrence of anticancer drugs in the aquatic environment: A systematic review. *Environmental Science and Pollution Research*, *27*(2), 1339–1347. https://doi.org/10.1007/s11356-019-07045-2

Ngqwala, N. P., & Muchesa, P. (2020). Occurrence of pharmaceuticals in aquatic environments: A review and potential impacts in South Africa. South African Journal of Science, 116(7–8), 42–48. https://doi.org/10.17159/sajs.2020/5730

Olarinmoye, O., Bakare, A., Ugwumba, O., & Hein, A. (2016). Quantification of pharmaceutical residues in wastewater impacted surface waters and sewage sludge from Lagos, Nigeria. Journal of Environmental Chemistry and Ecotoxicology, 8(3), 14-24.

Petrie, B., & Camacho-Munoz, D. (2020). Analysis, fate and toxicity of chiral non-steroidal anti-inflammatory drugs in wastewaters and the environment: A review. Environmental Chemistry Letters. https://doi.org/10.1007/s10311-020-01065-y

Valdez-Carrillo, M., Abrell, L., Ramirez-Hernandez, J., Reyes-Lopez, J. A., & Carreon-Diazconti, C. (2020). Pharmaceuticals as emerging contaminants in the aquatic environment of Latin America: A review. Environmental Science and Pollution Research, 27(36), 44863–44891. https://doi.org/10.1007/s11356-020-10842-9

Manuals:

Brian Ripley and Michael Lapsley (2020). RODATABASEC: ODATABASEC Database Access. R package version 1.3-17. https://CRAN.R-project.org/package=RODATABASEC

DIN EN ISO 3166-1. Codes für die Namen von Ländern und deren Untereinheiten - Teil 1: Codes für Ländernamen (ISO 3166-1:2013); Deutsche Fassung EN ISO 3166-1:2014

Jeroen Ooms, David James, Saikat DebRoy, Hadley Wickham and Jeffrey Horner (2020). RMySQL: Database Interface and 'MySQL' Driver for R. R package version 0.10.21. https://CRAN.R-project.org/package=RMySQL

Neuwirth, Erich (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. https://CRAN.R-project.org/package=RColorBrewer

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Roger Bivand and Colin Rundel (2020). rgeos: Interface to Geometry Engine - Open Source ('GEOS'). R package version 0.5-5. https://CRAN.R-project.org/package=rgeos

Roger Bivand and Nicholas Lewin-Koh (2021). maptools: Tools for Handling Spatial Objects. R package version 1.1-1. https://CRAN.R-project.org/package=maptools

Roger Bivand, Tim Keitt and Barry Rowlingson (2021). rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.5-23. https://CRAN.R-project.org/package=rgdal

A Appendix

Table 8: Number of positively detected MECs in surface water, groundwater or drinking water for substances occurring in all five UN-regions: WEOG – Western Europe and Others Group, GRULAC – Latin American and Caribbean Group, EEG – Eastern Europe Group, ASG – Asian Group, AFG – African Group

Substance	WEOG	EEG	ASG	AFG	GRULAC
Diclofenac	2214	546	261	34	145
Ibuprofen	2068	422	235	51	144
Carbamazepine	1726	456	377	68	108
Sulfamethoxazole	1176	177	816	84	53
Naproxen	626	364	176	26	120
Trimethoprim	562	13	490	43	47
Paracetamol	539	25	209	42	58
Gemfibrozil	498	67	120	1	39
Clarithromycin	367	12	308	17	9
Sulfamethazine	175	4	499	20	3
Bezafibrate	518	56	83	2	13
Ciprofloxacin	153	10	469	25	8
Atenolol	520	13	75	2	41
Estrone	265	74	188	51	60
Ketoprofen	287	206	107	4	18
Triclosan	284	39	204	6	59
17-alpha-Ethinylestradiol	285	12	99	3	87
Sulfadiazine	91	2	367	10	1
17-beta-Estradiol	228	28	116	49	102
Clofibric acid	278	17	112	2	5
Tetracycline	75	6	277	10	3
Iopromide	205	2	66	3	6
Metformin	206	2	39	2	4
Sulfathiazole	65	1	184	1	2
Venlafaxine	215	2	16	2	9

Substance	WEOG	EEG	ASG	AFG	GRULAC
Indometacin	110	69	53	1	10
Valsartan	163	12	43	2	5
Fluoxetine	131	2	67	2	1
Estriol	92	14	58	1	24
Triclocarban	58	1	117	5	4
Diazepam	121	21	34	3	2
Codeine	113	8	37	2	1
Metronidazole	54	1	81	12	5
Progesterone	73	2	12	6	26
10,11-Dihydro-10,11-Dihydroxy Carbamazepine	73	4	1	7	16
Atorvastatin	42	14	31	3	10
Acetylsalicylic acid	22	6	33	3	8