ZerOPM

## Zero pollution of Persistent, Mobile substances

Grant Agreement No. 101036756

## Deliverable 5.2

September 2025

# Circular Economy, Environmental, and Emissions database for the global inventory

## Work Package 5 – Substance Grouping

Deliverable Work Package Leader:
Swiss Federal Laboratories for Materials Science and Technology (EMPA)

Revision: 0
Dissemination level: Public

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101036756.

Any dissemination of results must indicate that it reflects only the author's view and that the Agency is not responsible for any use that may be made of the information it contains.

The present document has not yet received final approval from the European Commission and may be subject to changes.

## Note about contributors

| | |
|---|---|
| Lead partner responsible for the deliverable: | Swiss Federal Laboratories for Materials Science and Technology (EMPA) |
| Deliverable prepared by: | Zhanyun Wang, Huadong Xing, Raoul Wolf, Hans Peter Arp |
| Partner responsible for quality control: | Norwegian Geotechnical Institute AS |
| Deliverable reviewed by: | Emma Knight |
| Other contributors: | ETH Zürich, Jonathan Blumenthal;<br>NGI AS, Laura Monclus, Andrea Gredelj<br>DVGW-TZW, Sarah Hale |

## Project information

| | |
|---|---|
| Project period: | 01.10.2021 – 30.09.2026 |
| Duration (no. of months): | |
| Web-site: | www.zeropm.eu |
| Project coordinator: | Norwegian Geotechnical Institute AS, (NGI AS project no.: 20210423) |

## Project partners:

# Summary

The aim of Work Package 5 (WP5) Substance Grouping is to address the hypothesis that grouping persistent and mobile (PM) substances based on specific molecular substructures will support prioritization and directed risk assessment procedures of persistent and mobile substance groups. As part of this work the previous deliverable, D5.1, presented a new Global Chemical Inventory (GCI) that contained as many high-quality chemical structure information as possible; further, this also presented the outcome of in silico assessments of persistence and mobility, based on the classifications for persistent, mobile and toxic (PMT) and very persistent, very mobile (vPvM) substances in the recently amended Chemicals, Labelling and Packaging (CLP) regulation (EC 1272/2008).

The goal of this ZeroPM Deliverable, D5.2, is to categorize as many chemicals in the GCI as possible into categories of emission to aquatic environments, namely "***emissions confirmed***," "***emissions likely***," and "***no emission data available***." The following sections will begin with an overview of the various datasets used, followed by the results and discussion of the emission categories assigned to chemicals in the GCI. The document concludes with a brief discussion of limitations and a future outlook.

The data sources for this deliverable include the following inventories for specific chemical *uses* and inventories of chemical *monitoring data*:

- The chemical inventories for specific uses include databases from Canada, the US, the EU, and China, each contributing to the identification of chemicals likely to result in emissions. These inventories cover pesticides (PPID, APPRIL, EU Pesticide Active Substances), cosmetics (CosIng, IECIC), and biocides (EU Biocidal Active Substances). Collectively, thousands of chemicals from these sources—ranging from 4 in IECIC to over 2,500 in CosIng—are also found in the Global Chemical Inventory. Despite some substances potentially being inactive, all are categorized as "emissions likely" due to their intended applications and regulatory contexts.
- The inventories of chemical monitoring data for the aquatic environment include four major datasets that provide evidence of chemical presence in various environmental media. Two of which were expanded/updated within the work of ZeroPM, these are LitChemPlast and Arp and Hale (2022). The LitChemPlast database compiles data from 372 studies on chemicals analytically detected in plastics, identifying 1,432 substances also present in the Global Chemical Inventory and categorized as "emissions likely" due to the abundance of (micro)plastic pollution in aquatic resources. Arp and Hale (2022) reviewed 55 studies on organic chemicals in aquatic environments, identifying 873 overlapping substances, which are classified as "emissions confirmed. Two data sets external to ZeroPM are the Multimedia Monitoring Database (MMDB), which aggregates over 63 million records from multiple sources, with 1,398 chemicals matching the Global Chemical Inventory for substances monitored in the aquatic environment and therefore were considered "emissions confirmed." Lastly, the ANST/POL list from Muir et al. (2023) includes 8,112 chemicals

detected in environmental media over 50 years, categorized as “emissions likely”, as it was unclear if they were reported in the aquatic environment or another environmental media.

By compiling all datasets, a total of 10,556 substances from the Global Chemical Inventory were assigned emission categories to aquatic environments. 7,352 substances appeared in only one dataset, while 3,335 were found in multiple datasets, highlighting overlaps and reinforcing the value of integrating diverse sources. Specifically, 1,749 substances were classified as “emissions confirmed” based on direct environmental measurements, and 8,938 as “emissions likely” due to their intended uses or presence in relevant inventories. This distribution underscores the importance of using multiple datasets to ensure comprehensive and robust chemical emission assessments.

This database marks a significant initial effort to categorize the emission potential of chemicals listed in the Global Chemical Inventory by integrating seven datasets, including regulatory inventories and environmental measurements. While it covers over 10,500 substances, it still represents only a portion of the full inventory due to limitations in chemical use data, product measurements, and environmental monitoring. Challenges such as inconsistent terminology, limited coverage of non-plastic materials, and fragmented literature data hinder broader inclusion. Additionally, discrepancies arise because the Global Chemical Inventory focuses on industrial chemicals and UVCBs, whereas measurement datasets often include discrete substances, analogs, and transformation products. Despite these limitations, the database provides a valuable foundation for future expansion and refinement, aiming to support more comprehensive chemical emission assessments and sustainable circular economy practices, and to prioritize the persistent and mobile substances needing the most attention.

The data generated from this deliverable is available on Zenodo via : https://zenodo.org/records/17211524 and will be ultimately available as part of the searchable ZeroPM database (https://database.zeropm.eu/).

# Contents



# Tables



# Figure



# Appendices

# 1 Introduction

This work is part of the EU Horizon 2020 project, *Zero pollution of persistent, mobile substances (ZeroPM)*. ZeroPM integrates and synergizes three strategic approaches to protect environmental and human health from persistent and mobile substances: *Prevent*, *Prioritize*, and *Remove*. To achieve this, ZeroPM is developing an evidence-based, multilevel framework that will guide policy development, technological innovation, and market incentives to minimize the use, emissions, and pollution of entire groups of PM substances.

The aim of Work Package 5 (WP5) Substance Grouping is to address the central hypothesis that grouping persistent and mobile substances based on specific molecular substructures will support persistent and mobile substance prioritization and directed risk assessment procedures. Persistent and mobile substances are increasingly recognized as significant threats to human health and the enviroment. They not only persist in the environment, leading to long-term exposure, but their high mobility allows them to travel far from their sources, resulting in widespread contamination. In many cases, drinking water supplies must be treated using advanced, costly, and energy-intensive technologies to remove persistent and mobile substances in order to reduce human exposure. While many per- and polyfluoroalkyl substances (PFASs) are well-known examples, numerous other persistent and mobile substances also pose serious concerns. This global issue has prompted new research, monitoring efforts, and policy initiatives aimed at understanding and addressing persistent and mobile substances.

Through the seven tasks of WP5 in ZeroPM, WP5 will contribute to “safe and sustainable by design” strategies and the development of an “early warning system” for persistent and mobile substances, intended for use by regulators, industry, scientists, and water resource managers (Figure 1).

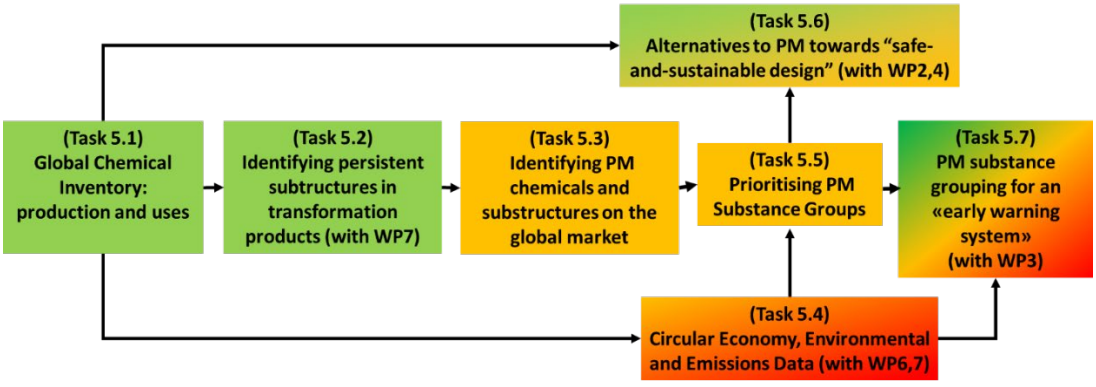


*Figure 1. Workflow of the tasks in WP5 of ZeroPM to prioritize PM substance groups for further action for regulatory authorities and towards safe and sustainable by design.*

This deliverable, which presents a circular economy, environmental, and emissions database, synthesizes the outcomes of Tasks 5.1 and 5.4. It builds upon the Global Chemical Inventory (GCI) developed in Task 5.1 and presented in Deliverable 5.1, serving as a foundation for integrating multiple chemical inventories for specific uses

such as pesticides and personal care products compiled under Task 5.1, along with inventories of chemicals detected in plastics and environmental measurements compiled and collected through Task 5.4.

The goal of this Deliverable 5.2 is to categorize as many chemicals in the GCI as possible into categories of emission to aquatic environments, namely "***emissions confirmed***," "***emissions likely***," and "***no emission data available***." The following sections will begin with an overview of the various datasets used, followed by the results and discussion of the emission categories assigned to chemicals in the GCI. The document concludes with a brief discussion of limitations and future outlook.

# 2 Overview of the data sources

For this deliverable, three types of datasets were considered: (1) the Global Chemical Inventory from Task 5.1, (2) chemical inventories for specific uses, and (3) inventories of chemical measurements. While the Global Chemical Inventory serves as the foundational dataset, the latter two provide supporting evidence for assigning emission categories to chemicals. The following subchapters describes these datasets in detail.

## 2.1 Global Chemical Inventory

The Global Chemical Inventory, originally developed by Wang et al. (2020), has been significantly enhanced under Task 5.1 of the ZeroPM project. This updated version integrates new data sources, a refined methodology for chemical structure identification, and a modern database infrastructure capable of supporting future expansions. Built on the Findable, Accessible, Interoperable, and Reproducible (FAIR) principles, the Global Chemical Inventory uses a code-first approach with open-source tools and data repositories, and adopts the InChI identifier for consistent chemical representation.

In brief, 16 out of 22 previously used data sources were verified and updated, covering 38 countries and resulting in over half a million unique entries. Data extraction involved handling diverse formats and applying extensive cleaning procedures to ensure compatibility. Chemical names and registry numbers were matched to structures using automated API services, and a consensus ranking method was applied to select the most probable structure for each entry. Ultimately, 144,351 unique chemical structures (a 3% increase compared to the work by Wang et al. (2020)), including 42,901 multicomponent entries, were identified and assigned ZeroPM IDs for further analysis.

## 2.2 Chemical inventories for specific uses

### 2.2.1 Canadian Pesticide Product Information Database (PPID)[1]

The PPID, maintained by Health Canada, provides public access to information on products, active ingredients, and programs related to pesticides and other pest control

[1] https://open.canada.ca/data/en/dataset/e10b0d6e-04ac-4014-a64a-666c3874bbe0

products regulated by the Pest Management Regulatory Agency. A total of 646 chemicals listed in the PPID are also found in the Global Chemical Inventory. These chemicals are assumed to fall under the category of "emissions likely," as it is acknowledged that some may no longer be in active use.

### 2.2.2 US Active Pesticide Product Registration Informational Listing (APPRIL)[2]

The APPRIL online database provides detailed information on active pesticide registrations and Supplemental Registrations (distributor products) approved by the US Environmental Protection Agency. Users can search the database using various criteria, including registration numbers, product names, company details, active ingredients, registration dates, and targeted sites or pests. The database is updated daily to ensure current information. A total of 1,290 chemicals listed in APPRIL are also found in the Global Chemical Inventory. These chemicals are assumed to fall under the category of "emissions likely," as it is recognized that some may no longer be in active use.

### 2.2.3 EU Cosmetics Ingredients (CosIng) Database[3]

CosIng is the European Commission's database for information on cosmetic substances and ingredients, as outlined in the Cosmetics Regulation (EC) No 1223/2009, the Cosmetics Directive 76/768/EEC (as amended), the Glossary of common ingredient names for cosmetic product labelling (established by Decision (EU) 2019/701), and the Scientific Committee for Consumer Safety (SCCS) opinions on cosmetic ingredients. The database allows searches by CAS, ELINCS, or EINECS numbers and includes data dating back to the adoption of the Cosmetics Directive in 1976. Current entries are marked as "active," while historical ones are listed as "not active." A total of 2,515 chemicals listed in CosIng are also found in the Global Chemical Inventory. These chemicals are assumed to fall under the category of "emissions likely," as it is acknowledged that some may no longer be in active use.

### 2.2.4 Inventory of Existing Cosmetic Ingredients in China (IECIC)[4]

IECIC 2021, issued by the Chinese National Medical Products Administration (NMPA), is a comprehensive list of existing cosmetic ingredients that have been used in cosmetics in China. Ingredients not listed in IECIC are considered new and require registration or filing before use. The database includes a total of 8,972 ingredients. Of these, 4 chemicals are also found in the Global Chemical Inventory. These chemicals are assumed to fall under the category of "emissions likely," as it is acknowledged that some may no longer be in active use.

### 2.2.5 EU Pesticide Active Substances[5]

[2] https://ordspub.epa.gov/ords/pesticides/f?p=APPRIL_PUBLIC:2::::::
[3] https://ec.europa.eu/growth/tools-databases/cosing/
[4] https://hzpsys.nifdc.org.cn/hzpGS/ysyhzpylml#
[5] https://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/start/screen/active-substances

This database provides information on active substances, including those classified as low-risk or candidates for substitution, as well as basic substances, whether approved or not within the EU. It also includes certain safeners and synergists, although these have not yet undergone assessment at the EU level. A total of 219 chemicals listed in the EU Pesticide Active Substances are also found in the Global Chemical Inventory. These chemicals are assumed to fall under the category of "emissions likely," as it is acknowledged that some may no longer be in active use. For the purposes of this database, only substances currently listed as "approved" are considered in the core dataset. Supplementary information on non-approved substances is included for reference and to support further exploration by interested users.

### 2.2.6 EU Biocidal Active Substances[6]

Under the Biocidal Products Regulation ((EU) No 528/2012, BPR), a biocidal product must be authorised before it can be placed on the market or used within the European Economic Area (EEA) and Switzerland. This authorisation process involves two sequential steps. First, the active substance is evaluated and, if it meets the required criteria, is approved for use in a specific product type. The second step involves the authorisation of each biocidal product that contains, consists of, or generates the approved active substance(s). This database pertains to the first step of the process. It includes active substances that are part of the Review Programme, those outside the Review Programme, and substances listed in Annex I of the BPR. A total of 19 chemicals listed in the EU Biocidal Active Substances are also found in the Global Chemical Inventory. These chemicals are assumed to fall under the category of "emissions likely," as it is acknowledged that some may no longer be in active use.

## 2.3 Inventories of Chemical Measurements

### 2.3.1 LitChemPlast Database

Under Task 5.4 of the ZeroPM project, a scoping review was conducted to compile studies measuring chemical content in plastics into a new open-access LitChemPlast database. Chemicals in plastics were used as a proxy for chemicals in the circular economy, due to the widespread use of plastics in various materials, and the ambition to reuse/recycle more plastics within the context of the circular economy.

The review conducted to develop the LitChemPlast section of the database presented here included studies from 1978 to 2021 that analyzed raw plastic materials, products (e.g., packaging, construction, agriculture), and waste. Studies were identified through keyword searches in Web of Science and SciFinder, followed by a two-stage screening process. Although food packaging and textiles were not specifically targeted, they were included when encountered. Data extracted from each study included bibliographic details, sample types, analytical methods, and chemical concentrations. To ensure consistency, chemicals were assigned CAS numbers and categorized by substance group, product type, polymer type, sampling stage, and recycling status. Analytical

[6] https://www.echa.europa.eu/information-on-chemicals/biocidal-active-substances

methods were classified as either targeted or non-targeted, and concentration data were standardized to mg/kg wherever possible.

The final LitChemPlast database includes data from 372 studies from 786 studies, identifying over 3,500 substances across various stages of the plastic life cycle (in total, more than 47,000 samples have been tested and over 65,000 entries in the LitChemPlast). Notably, around 75% of these substances were detected only through non-targeted screening, while targeted analyses focused on well-known groups such as metal(loid)s, brominated flame retardants, and *ortho*-phthalates. Some economically significant product categories—such as non-food packaging, construction, and automotive plastics—remain underrepresented in the literature. Recycled plastics were also less frequently studied, though existing data suggest they contain higher levels of regulated flame retardants. The LitChemPlast database offers a valuable resource for exposure assessments and substance flow analyses, highlighting the need for more comprehensive data to support a safe circular plastics economy. A deeper discussion of this database has been recently published (Wiesinger et al., 2024).

While the information of the LitChemPlast database is plentiful, for this deliverable only relevant information was extracted: product group(s), product details, recycling information, sampling information, country information, CAS Registry Numbers, InChIKeys, detection frequencies, and the maximum concentration (ppm). Only entries with either a detection frequency above 0 and/or a maximum concentration above 0 were considered. A total of 1,432 chemicals listed in LitChemPlast are also found in the Global Chemical Inventory. These chemicals are assumed to fall under the category of "emissions likely," as it is acknowledged that some may no longer be in active use.

### 2.3.2 Detected Substances in Freshwater by Arp and Hale (2022)

In Arp and Hale (2022), a literature review was conducted to compile monitoring studies of organic chemicals detected in six types of aquatic media: wastewater treatment plant effluent, surface water, bank filtrate, groundwater, raw water, and drinking water. The search, covering the years 2000–2019, used Google Scholar and Web of Science with keywords such as "organic chemical" and "contaminant," without filtering for region or treatment methods. The goal was not to be exhaustive but to gather a sufficiently large dataset to analyze the persistence and mobility of detected substances. In total, 55 unique studies or compilations were included, many covering multiple water types.

Across these studies, 1,289 unique organic chemicals were identified. Surface water had the highest number of unique detections (1,021), largely due to the availability of comprehensive compilations. In contrast, bank filtrate and raw water had the fewest unique detections (114 and 212, respectively), reflecting the limited number of studies available. Groundwater and drinking water had 338 and 385 unique substances detected, respectively. This dataset includes both target analysis and non-target analysis, and provides a foundation for understanding the environmental distribution of organic contaminants across different stages of the water cycle.

From this data collection, information on the media, countries, CAS Registry Numbers, and maximum concentration (ng/L) were extracted. The CAS Registry Numbers were verified using the check-digit method mentioned above and used to retrieve PubChem CID identifiers using PubChem's PUG REST API services (https://pubchem.ncbi.nlm.nih.gov/docs/pug-rest). The PubChem CIDs were then used to retrieve structural information on the substances in the form of InChIKeys. A total of 873 chemicals listed in Arp and Hale (2022) are also found in the Global Chemical Inventory. These chemicals are assumed to fall under the category of "emissions confirmed," as these are environmental measurements. Some of the substances not present on the Global Chemical Inventory were transformation products and pesticides, identified through non-target analysis,

### 2.3.3 Multimedia monitoring database (MMDB)[7]

MMDB, published in 2022 by Isaacs et al., was developed by compiling data from several reputable monitoring databases: American Healthy Homes Survey, National Atmospheric Deposition Program, Biomonitoring California, California Air Monitoring Network, California Surface Water Database, California Air Resources Board (CARB), ChemTheatre, Comparative Toxicogenomics Database, EPA Nine POTW Study, US EPA Ambient Monitoring Technology Information Center – Air Toxics Data, US EPA Discharge Monitoring Report, US EPA Office of Water – National Study of Chemical Residues in Lake Fish Tissue, Targeted National Sewage Sludge Survey, US EPA Unregulated Contaminant Monitoring Rule, US FDA total Diet Study, ICES-DOME, EU Information Platform for Chemical Monitoring Data (IPCHEM), US National Health and Nutrition Examination Survey, USDA National Residue Program (NRP) and USGS Monitoring Data – National Water Quality Monitoring Council. The compilation process involved both automated and manual curation techniques to ensure the accuracy and reliability of the information.

The database contains 63,768,583 harmonized data records, comprising 54,520,407 single-sample records and 9,248,176 aggregate records. From these, 9,956 unique raw chemical identifiers (based on names and/or CAS Registry Numbers) were identified. Of these, 8,757 substances could be mapped to 3,271 unique DTXSIDs in the DSSTox database. The mapped chemicals span a broad range of substance and use classes, including metals, pesticides, flame retardants, polychlorinated biphenyls (PCBs), pharmaceuticals, and chemicals used in both consumer and industrial applications.

Due to the large size of the database (>20 GB in compressed format) and absence of any user interface, we did not include all information from the database, but only the lists of substances that have been detected in various aquatic compartments, including information on the DTXSID, CAS Registry Number, and InChIKey. A total of 1,398 chemicals listed in MMDB are also found in the Global Chemical Inventory. These chemicals are assumed to fall under the category of "emissions confirmed," as these are environmental measurements.

[7] https://epa.figshare.com/articles/dataset/Multimedia_Monitoring_Database_MMDB_/17065024?file=31555780

A central difference between MMDB and Arp and Hale (2022) was that the monitoring data in MMDB goes further back in time and does not include more recent data from non-target analysis, whereas the Arp and Hale (2022) data set includes monitoring data up to the year 2000 and included more recent data from non-target analysis.

### 2.3.4 List of Substances Analyzed in Environmental Media (ANST/POL)

Muir et al. (2023) conducted a bibliometric survey to identify which chemicals have been detected in environmental media and to analyze their trends over the past 50 years. The study utilized the CAplus database from CAS (a division of the American Chemical Society), searching for publications indexed with the roles “analytical study” and “pollutant” from 1978 to 2018. These indexing roles were considered comprehensive enough to capture substances identified in environmental media, while excluding toxicology studies, chemical synthesis, and reagent use. Patents were not included.

CAS Registry Numbers (CASRNs) associated with these roles were compiled annually, resulting in many duplicates due to repeated reporting of substances across years. A final, non-redundant list, the ANST/POL list, was created by sequentially removing duplicates, starting from the earliest year, using a duplicate finder tool. This process was completed in April 2018.

The final ANST/POL list comprised 19,776 unique CASRNs, with a total of 105,410 citations for the period January 1978 to April 2018. For each CASRN, InChI and SMILES codes, as well as chemical names, were generated. Approximately 900 CASRNs could not be structurally characterized through database searches, as they represented mixtures, trade names, polymers, proteins, or otherwise undefined substances; these were manually reviewed using PubChem and SciFinder. For about 300 partially defined substances, representative SMILES and InChI codes were assigned by drawing likely structures (e.g., for bromo- or methoxy-diphenyl ethers, bromo-chloro carbazoles, chloroalkanes, alkylated PAHs, PCBs as Aroclors, Phenoclor, Kanechlors, or polybrominated diphenyl ethers as Bromkal).

Of the identified substances, about 14,150 CASRNs were found on various priority lists or were close analogs and transformation products. Notably, substances listed in the industrial chemical inventories of Europe, China, and the United States accounted for only about 5% of the measured substances. In contrast, pharmaceuticals and current-use pesticides were widely detected, representing 50–60% of total CASRN counts between 2000 and 2015. A total of 8,112 chemicals listed in ANST/POL list are also found in the Global Chemical Inventory. These chemicals are assumed to fall under the category of “emissions likely,” as these are environmental measurements for all types of media, which may or may not be relevant for aquatic environments.

## 2.4 Overview of the Circular Economy, Environmental, and Emissions Database

When compiling all datasets together, a total of 10,556 substances from the Global Chemical Inventory were found to be covered (seeTable 1). Among these, 7,352 substances appeared in only one dataset, while 3,335 substances were found in two or more datasets. Specifically, 1,947 substances appeared in two datasets, 735 in three, 353 in four, 185 in five, 87 in six, 26 in seven, and 2 in eight datasets considered. Examples of the overlaps of substances across the data sources are illustrated in Figure 2 and Figure 3.

This distribution highlights the importance of using multiple data sources to assess the emission potential of chemicals. Each dataset offers a unique but limited scope, and relying on a single source may overlook substances that are relevant in other contexts. Integrating diverse datasets enhances the robustness of chemical assessments and supports more comprehensive environmental evaluations.

These 10,556 substances in the Global Chemical Inventory have been assigned categories of emissions to aquatic environments, with 1,749 substances labelled as "emissions confirmed" and 8,938 substances labelled as "emissions likely."

*Table 1. Overview of the Circular Economy, Environmental, and Emissions Database.*

| Data source | Number of substances in the Global Chemical Inventory | (Quantification) information contained in the data source | Media/ Environmental Compartments | Spatio-temporal information |
|---|---|---|---|---|
| ***Chemical inventories for specific uses*** | | | | |
| Canadian Pesticide Product Information Database (PPID) | 646 | Reevaluation status | N.A. | Canada |
| US Active Pesticide Product Registration Informational Listing (APPRIL) | 1290 | Registration status, pesticide type, pesticide category, use type, active ingredients, sites, pests, etc. | N.A. | US |
| EU Cosmetics Ingredients (CosIng) Database | 2515 | INCI name, description, functions, etc. | N.A. | EU |
| Inventory of Existing Cosmetic Ingredients in China (IECIC) | 4 | Application site, maximum historical usage (%), safety and technical standards for cosmetics, etc. | N.A. | China |
| EU Pesticide Active Substances | 219 | Approval status, expiration of approval, legislation, remark, risks, candidate for substitution, ADI, ARfD, AOEL, etc. | N.A. | EU |
| EU Biocidal Active Substances | 19 | Approval status, approval end date, product type, review programme, new active | N.A. | EU |

| | | | | |
|---|---|---|---|---|
| | | substance, annex I substance, assessment status, legal act, etc. | | |
| ***Inventories of Chemical Measurements*** | | | | |
| LitChemPlast (Wiesinger *et al.*, 2024) | 1432 | Detection frequency (where available), concentration (where available) | 192 product groups, 1,641 detailed products, recycled products | 70 countries, years 1968–2022 |
| Arp & Hale (2022) | 873 | Presence, concentration (where available) | Bank filtrate, drinking water, ground water, surface water, wastewater treatment plant effluent | 56 countries |
| Multimedia monitoring database (MMDB; Isaacs *et al.*, 2022) | 1398 | Presence (see individual description) | Drinking water, ground water, human blood, sediments, sludge, surface water, urine, wastewater | None (see individual description) |
| List of Substances Analyzed in Environmental Media (ANST/POL; Muir et al., 2024) | 8112 | Number of studies | All environmental media | None (see individual description) |

N.A. = not applicable

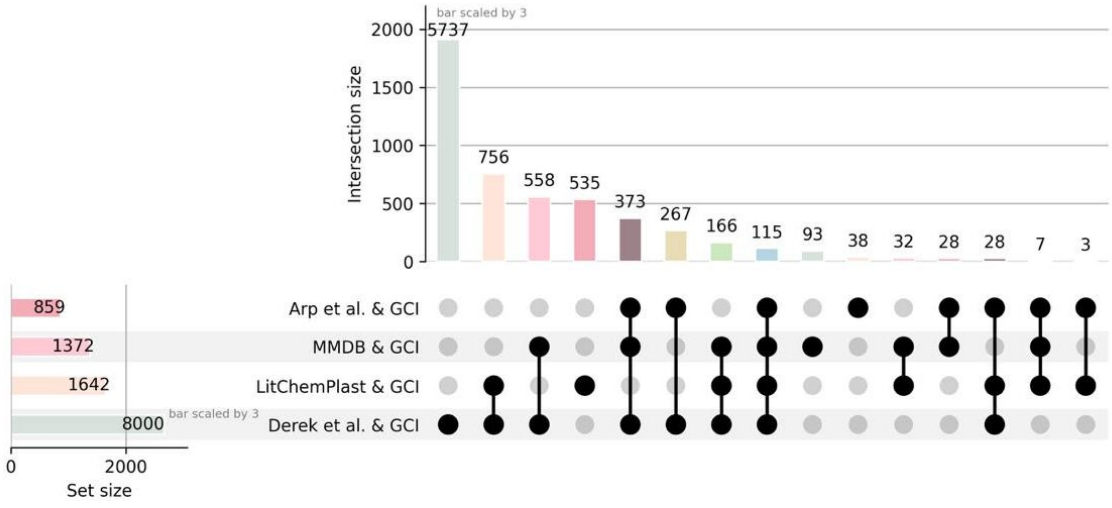


*Figure 2. Overview of the number of chemical substances in the four inventories of chemical measurements that are also in the Global Chemicals Inventory. It shows the overlap of four sets, each representing the interaction of Global Chemicals Inventory with corresponding dataset. The left bars indicate the number of each set, and the top bars show the intersection sizes. The black points in the matrix below mark which sets contribute to each intersection. To improve readability, the largest set and the largest intersection were scaled down by 3 for display (annotated as "bar scaled by 3"), while the labels report the actual counts.*

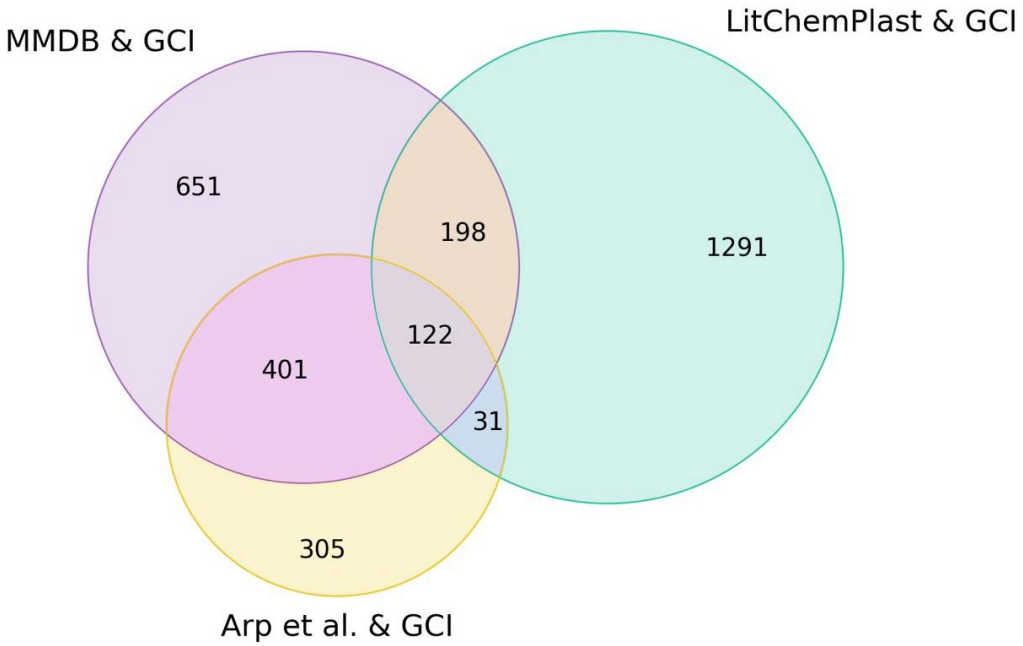


*Figure 3. Venn diagram showing the overlap of substances among three sets. Each set represents the interaction between the Global Chemicals Inventory and the corresponding dataset. The diagram also highlights the substances that are exclusive to each respective data source. The Venn diagram circles is proportional to the number of substances they represent.*

## 2.5 Limitations and Outlook

This database represents an important first step toward categorizing the emission potential of chemicals listed in the Global Chemical Inventory. It draws on seven datasets, including inventories of chemicals for specific uses and measurements of chemicals in products and the environment. One dataset, developed under Task 5.4, involved reviewing over 700 publications and manually extracting more than 65,000 data entries. The remaining six datasets were sourced from various regulatory and scientific efforts, with at least two targeting specifically detections in aquatic media, Arp and Hale (2022) and MMDB, representing substantial prior work and which summarized earlier efforts of compiling detection data. Despite the extensive effort and the large number of chemicals labeled (over 10,500 substances), the database still covers only a portion of the substances in the Global Chemical Inventory. The present database has the following limitations:

1. The scope of chemical use information. The current database primarily builds upon inventories of chemicals for specific uses, particularly pesticides and cosmetics. Meanwhile, additional chemical use data are available in other comprehensive databases, such as the ECHA's REACH registration database, the Substances in Preparations in Nordic Countries (SPIN) database, and the US EPA's Chemical Data Reporting (CDR) database. These sources were not included in the present study due to

the complexity involved in machine-processing the use information they contain. This complexity arises from several factors, including: ambiguous use terms that make it difficult to interpret their relevance to emission potential to aquatic environments, conflicting use information for the same chemical across different data entries in the same inventory, and inconsistencies in terminology both across databases and over time within the same database. These challenges highlight the need for future efforts to systematically explore how chemical use reporting can be standardized and made fit-for-purpose for environmental assessments and others. A commentary article is currently being prepared to raise awareness of this issue and encourage broader consideration within the scientific and regulatory communities.

2. The scope of measurements of chemicals in products. The current study focuses on the presence of chemicals in plastics, representing the circular economy component of the database. While plastics are a major material undergoing circular processes, other materials, such as paper-based products, also participate in circular systems. However, compared to plastics, significantly fewer studies have investigated the chemical content of paper materials. Although plastics remain a critical focus for further investigation, future research should expand to include other material types within circular systems. This will help build a more comprehensive understanding of chemical emissions and support the development of safer and more sustainable circular economy practices.

3. The scope of measurements of chemicals in the environment. While the present study includes two large datasets, additional measurement data are available in the literature. However, compiling such data remains a daunting task, as they are dispersed across numerous publications in the public domain and presented in varying formats, often not in machine-readable form. Although the adoption of Open and FAIR principles is growing, extracting non-structured data from literature continues to pose challenges. Recent advancements in large language models may offer new opportunities to streamline this process, but their practical application for extracting chemical measurement data from scientific literature remains to be fully explored. Future efforts should investigate how these technologies can be leveraged to improve data accessibility and integration in environmental research.

Furthermore, it is observed that a significant number of chemicals identified in the consulted datasets are not present in the Global Chemicals Inventory. Several factors contribute to this divergence. For example, the Global Chemicals Inventory primarily focuses on industrial chemicals and generally excludes chemicals that are specifically used for certain categories such as pesticides, biocides, and pharmaceuticals (though the Global Chemicals Inventory includes multiple-purpose chemicals). Also, the Global Chemicals Inventory encompasses substances of unknown or variable composition, complex reaction products, or biological materials (UVCBs), which are often complex mixtures. In contrast, measurement inventories typically list only discrete chemicals, some of which may be components of UVCBs. However, due to limited information on the composition of UVCBs, it is not possible to fully map these overlaps. Furthermore, the Global Chemicals Inventory lists intentionally added chemicals, whereas measurement inventories may also include close analogs and transformation products, as highlighted by Derek et al. (2024). Arp and Hale (2022) include several

transformation products that were not included, as these were not intentionally added chemicals to the global chemical market. These discrepancies underscore the importance of continued efforts to expand and refine the Global Chemicals Inventory.

Despite these limitation, it is our intent that that by making the current circular economy, environmental, and emissions database available to the broader scientific and regulatory communities, it will inspire further efforts to expand and enhance the database, ultimately aiming to achieve comprehensive coverage of all chemicals listed in the Global Chemical Inventory.

# 3 Next steps towards Prioritization and Substance Grouping

Within ZeroPM, we are developing methodologies for the prevention, prioritization and removal of persistent and mobile substances, as well as persisetant and mobile substance groups. Key steps that have been presented in the previous deliverable 5.1 were i) identifying, to the best of our ability, the actual structural information of substances on the GCI; ii) assessing their persistence and mobility using a diverse array of experimental data and models that were integrated through Bayesian analysis, and iii) developing methodologies to identify dead-end transformation products of substances on the GCI and determine whether they would be persistence and mobility. These deliverable added new information and that was iv) to build a circular economy, environmental and emissions so that it is known for which of the persistent and mobile substances there may be emissions and therefore an indicator of risk for persistent and mobile substances. The next steps will be on the two following deliverables (Table 2).

*Table 2. Upcoming Deliverables in WP5*

| Deliverables | Description |
|---|---|
| **D5.3 (July 2026)** | Prioritized PM Substance and Substance Group suspect list, and non-PM alternatives for action by policy makers and researchers |
| **D5.4 (July 2026)** | Policy oriented technical and practical guideline for an "early-warning system" for priority PM substances groups |

The future work in D5.3 is to develop a prioritized persistent and mobile substance and substance group suspect list, and non-PM alternatives. For this purpose, we will use substances that are persistent and mobile in the GCI (task 5.1) or lead to a persistent and mobile transformation products (as identified in Task 5.2), and are associated with emissions (5.2) and toxicity (if available), to identify prioritized persistent and mobile substance groups. In collaboration with other activities in ZeroPM (primarily in WP2 and WP4), we will also identify potential non-persistent and mobile alternatives for their replacement.

Further, in D5.4 we will present a list of substances and substructures for water monitoring, as a suspect list for researchers and water utilities /regulators (based on tools from Task 5.4), that could be coupled with non-target analysis methods.

# 4 Supplementary Data Files

The output of this deliverable is presented on Zenodo via https://zenodo.org/records/17211524. More details about the data presented in this file can be found in the Appendix.

After approval of this deliverable, it will also be available on the searchable ZeroPM database (https://database.zeropm.eu/). Publication of this work on the ZeroPM database will allow users to identify which PM substances have been registered on the GCI and within a specific country or region. This database, building on the FAIR principles, will enable the broader regulatory and research community to develop their own approaches to prioritize PM substance groups in their specific country

# 5 References

Arp HPH, Hale S. 2022. Assessing the persistence and mobility of organic substances to protect freshwater resources. *ACS Environmental Au* 2(6), 482-509.

Isaacs KK, Wall JT, Williams AR, Hobbie KA, Sobus JR, Ulrich E, Lyons D, Dionisio KL, Williams AJ, Grulke C, Foster CA, McCoy J, Bevington C. 2022. A harmonized chemical monitoring database for support of exposure assessments. *Scientific Data* 9, 314.

Muir DCG, Getzinger GJ, McBride M, Ferguson PL. 2023. How many chemicals in commerce have been analyzed in environmental media? A 50 year bibliometric analysis. *Environmental Science & Technology* 57(25), 9119-9129.

Wang Z, Walker GW, Muir DCG, Nagatani-Yoshida K. 2020. Toward a global understanding of chemical pollution: a first comprehensive analysis of national and regional chemical inventories. *Environmental Science & Technology* 54(5), 2575-2584.

Wiesinger H, Shalin A, Huang X, Sigrist A, Plinke N, Hellweg S, Wang Z. 2024. LitChemPlast: an open database of chemicals measured in plastics. *Environmental Science & Technology Letters* 11(11), 1147-1160.

# Appendix. Overview of the data columns within the the ZeroPM circular economy, environmental, and emissions database

Within the supplementary data files, data is organized under the following column headings presented in Table A 1. Overview of the Circular Economy, Environmental, and Emissions Database.

*Table A 1. Overview of the Circular Economy, Environmental, and Emissions Database.*

| Column | Description | Source |
|---|---|---|
| Substance_SID | Substance ID in PubMed | ZeroPM: Global Chemicals Inventory (GCI) |
| Compound_CID | Compound ID in PubMed | |
| Data_Source | ZeroPM GCI as the data source | |
| External_ID | ZeroPM ID | |
| Synonyms | Synonym to ZeroPM ID | |
| Last_Modified_Date | Last modified date in GCI | |
| Deposit_Date | Deposit data in GCI | |
| Data_Source_Category | Data source category | |
| Iupac_name | IUPAC name | |
| Molecular_formula | Molecular formula | |
| Molecular_weight | Molecular weight | |
| Synonyms | Synonyms to IUPAC name | |
| Inchi | InChI (International Chemical Identifier) | |
| Inchikey | InChI key, hashed version of InChI | |
| Smiles | Simplified Molecular Input Line Entry System | |
| Cas | Chemical Abstracts Service Registry Numbers | |
| 0_Emissions_Confirmed | Chemicals identified as "emissions confirmed to aquatic environments" based on their presence in CAN_PPID, US_APPRIL, EU_CosIng, IECIC, EU_Pesticides, EU_Biocides, LitChem, Derek et al. | - |
| 0_Emissions_Likely | Chemicals identified as "emissions likely to aquatic environments" based on their presence in Arp & Hale and MMdB | - |
| 0_occurrence | The number of occurrence in the different databases | - |
| 1_1_CAN_PPID | Presence in the Canadian Pesticide Product Information Database (1=yes, 0=no) | https://open.canada.ca/data/en/dataset/e10b0d6e-04ac-4014-a64a-666c3874bbe0 |
| 1_2_US_APPRIL | Presence in the US Active Pesticide Product Registration Informational Listing (1=yes, 0=no) | https://ordspub.epa.gov/ords/pesticides/f?p=APPRIL_PUBLIC:2:::::: |

| Column | Description | Source |
|---|---|---|
| 1_3_EU_CosIng | Presence in the EU Cosmetics Ingredients (CosIng) Database (1=yes, 0=no) | https://ec.europa.eu/growth/tools-databases/cosing/ |
| 1_4_IECIC | Presence in the Inventory of Existing Cosmetic Ingredients in China (1=yes, 0=no) | https://hzpsys.nifdc.org.cn/hzpGS/ysyhzpylml# |
| 1_5_EU_Pesticides | Presence in the EU Pesticide Active Substances (1=yes, 0=no) | https://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/start/screen/active-substances |
| 1_6_EU_Biocides | Presence in the EU Biocidal Active Substances (1=yes, 0=no) | https://www.echa.europa.eu/information-on-chemicals/biocidal-active-substances |
| 2_1_LitChem | Presence in the LitChemPlast Database (1=yes, 0=no) | Wiesinger H, Shalin A, Huang X, Sigrist A, Plinke N, Hellweg S, Wang Z. 2024. LitChemPlast: an open database of chemicals measured in plastics. *Environmental Science & Technology Letters* 11(11), 1147-1160 |
| 2_2_Arp_Hale | Presence in the Detected Substances in Freshwater by Arp and Hale (2022) (1=yes, 0=no) | Arp HPH, Hale S. 2022. Assessing the persistence and mobility of organic substances to protect freshwater resources. *ACS Environmental Au* 2(6), 482-509 |
| 2_3_MMdB | Presence in the Multimedia monitoring database (1=yes, 0=no) | Isaacs KK, Wall JT, Williams AR, Hobbie KA, Sobus JR, Ulrich E, Lyons D, Dionisio KL, Williams AJ, Grulke C, Foster CA, McCoy J, Bevington C. 2022. A harmonized chemical monitoring database for support of exposure assessments. *Scientific Data* 9, 314 |
| 2_4_Derek_et_al | Presence in the List of Substances Analyzed in Environmental Media compiled by Derek et al. (2024) (1=yes, 0=no) | Muir DCG, Getzinger GJ, McBride M, Ferguson PL. 2023. How many chemicals in commerce have been analyzed in environmental media? A 50 year bibliometric analysis. *Environmental Science & Technology* 57(25), 9119-9129 |
| A_5_supporting_info_active_substance_id | Supporting information of the id in the database | EU Pesticide Active Substance Database |
| A_5_supporting_info_status_ec_1107_2009 | Supporting information of the approval status in the database | |

| Column | Description | Source |
|---|---|---|
| A_5_supporting_info_legislation | Supporting information of the relevant EU legislation in the database | |
| A_5_supporting_info_remark | Supporting information of the remarks in the database | |
| A_5_supporting_info_classification_clp_1272_2008 | Supporting information of the CLP classification in the database | |
| A_5_supporting_info_authorised | Supporting information of the authorisation status in the database | |
| A_5_supporting_info_functions | Supporting information of the functions in the database | |
| A_6_supporting_info_ec_list_no | Supporting information of the EC No. in the database | EU Biocidal Active Substance Database |
| A_6_supporting_info_bas_number | Supporting information of the BAS number in the database | |
| A_6_supporting_info_approval_status | Supporting information of the approval status in the database | |
| A_6_supporting_info_approval_end_date | Supporting information of the approval end date in the database | |
| A_6_supporting_info_product-type | Supporting information of the product type in the database | |
| A_6_supporting_info_review_programme_flag | Supporting information of whether the chemical being reviewed in the database | |
| A_6_supporting_info_new_active_substance_flag | Supporting information of whether the chemical being a new active substance in the database | |
| A_6_supporting_info_annex_i_substance_flag | Supporting information of whether the chemical being in the Annex I of the EU biocidal legislation in the database | |
| A_6_supporting_info_assessment_status | Supporting information of the assessment status in the database | |
| A_6_supporting_info_assessment_sub_status | Supporting information of the assessment sub-status in the database | |
| A_6_supporting_info_legal_act | Supporting information of the relevant legislation in the database | |
| A_6_supporting_info_link_to_legal_act | Supporting information of the link to the relevant legislaiton in the database | |
| A_6_supporting_info_candidate_for_substitution | Supporting information of whether the chemical being a candidate for substitution in the database | |
| A_6_supporting_info_factsheet_url | Supporting information of the link to a factsheet in the database | |
| A_6_supporting_info_related_authorised_biocidal_products | Supporting information of the authorisation status in the database | |
| A_6_supporting_info_substance_information_page | Supporting information of the substance information page in the database | |
| B_1_LitChem_supporting_info_product_group_level3 | Supporting information of the relevant plastic product group in the database | Wiesinger H, Shalin A, Huang X, Sigrist A, Plinke N, Hellweg S, Wang Z. |

| Column | Description | Source |
|---|---|---|
| B_1_LitChem_supporting_info_product_details_level4 | Supporting information of the relevant product details in the database | 2024. LitChemPlast: an open database of chemicals measured in plastics. *Environmental Science & Technology Letters* 11(11), 1147-1160. |
| B_1_LitChem_supporting_info_recycled | Supporting information of the sample being recycled or not in the database | |
| B_1_LitChem_supporting_info_sampling_stage | Supporting information of the sampling stage in the database | |
| B_1_LitChem_supporting_info_sampling_country | Supporting information of the sampling country in the database | |
| B_1_LitChem_supporting_info_freq_detection_percent | Supporting information of the detection frequency in the database | |
| B_1_LitChem_supporting_info_max_ppm | Supporting information of the maximum concentration detected [ppm] in the database | |
| B_2_Arp_Hale_supporting_info_environment_media | Supporting information of the environmental media detected in the database | Arp HPH, Hale S. 2022. Assessing the persistence and mobility of organic substances to protect freshwater resources. *ACS Environmental Au* 2(6), 482-509. |
| B_2_Arp_Hale_supporting_info_country | Supporting information of the sampling country in the database | |
| B_2_Arp_Hale_supporting_info_max_ng_l | Supporting information of the maximum concentration [ng/L] in the database | |
| B_3_MMdB_supporting_info_dtxsid | Supporting information of the relevant DTXSID according to the US EPA CompTox in the database | Isaacs KK, Wall JT, Williams AR, Hobbie KA, Sobus JR, Ulrich E, Lyons D, Dionisio KL, Williams AJ, Grulke C, Foster CA, McCoy J, Bevington C. 2022. A harmonized chemical monitoring database for support of exposure assessments. *Scientific Data* 9, 314. |
| B_3_MMdB_supporting_info_mmdb | ID in MMdB | |
| B_4_Derek_supporting_info_category | Supporting information of the chemical category assigned in the database | Muir DCG, Getzinger GJ, McBride M, Ferguson PL. 2023. How many chemicals in commerce have been analyzed in environmental media? A 50 year bibliometric analysis. *Environmental Science & Technology* 57(25), 9119-9129. |
| B_4_Derek_supporting_info_inventory_presence_EU_US_China | Supporting information of whether the chemical being listed in the EU, US or China in the database | |
| B_4_Derek_supporting_info_citation_counts | Supporting information of the number of citations assigned in the database | |

H2020 project Zero pollution of Persistent, Mobile substances
EU Grant Agreement No. 101036756