Texte
105/2022

Final report

PROSOIL – Protection of soil organisms: Development of toxicity criteria for soil organisms in the framework of classification of substances and PBT assessment

by:

Björn Scholz-Starke
darwin statistics, Aachen, Germany

Felix Stibany, Monika Hammers-Wirtz
Research Institute for Ecosystem Analysis and Assessment - gaiac, Aachen, Germany

publisher:
darwin statistics

German Environment Agency

TEXTE 105/2022

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

Project No. (FKZ) 3719 65 407 0

Report No. (UBA-FB) FB000843/ENG

Final report

PROSOIL – Protection of soil organisms: Development of toxicity criteria for soil organisms in the framework of classification of substances and PBT assessment

by

Björn Scholz-Starke
darwin statistics, Aachen, Germany

Felix Stibany, Monika Hammers-Wirtz
Research Institute for Ecosystem Analysis and Assessment - gaiac, Aachen, Germany

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
Internet: www.umweltbundesamt.de

/umweltbundesamt.de
/umweltbundesamt

Report performed by:
darwin statistics
Von-Broich-Str. 23
52072 Aachen

Report completed in:
March 2022

Edited by:
Enken Hassold (IV 2.3) and Korinna Ziegler (IV 1.2), Coordination Group "Ecotoxicology and Bioaccumulation" in Department IV

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

ISSN 1862-4804

Dessau-Roßlau, September 2022

Disclaimer

The views and opinions expressed in this report are those of the author(s) and do not necessarily reflect those of the German Environment Agency and must under no circumstances be regarded as stating an official position of the German Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection or the Federal Republic of Germany. The disclaimer comprises the responsibility for the methodologies used for data compilation, data harmonization, quality assurance, comparison with existing classifications, and interpretation of the results and conclusions.

The project aimed at deriving ecotoxicological sensitivity ranges for soil organisms as a scientific basis to derive suitable toxicity criteria for use in the context of classifications under the CLP regulation or for PBT assessments. Hence, this report may support ongoing discussions in the EU on terrestrial toxicity criteria. Please note that the proposed toxicity criteria for soil organisms and the analyses of the relevance and possible benefits are preliminary proposals of the author(s) and do not necessarily reflect the position of the editors and the German Environment Agency. Further analyses and discussions are needed to conclude legislative implementation based on the data and results available in this report.

Abstract: PROSOIL – Protection of soil organisms: Development of toxicity criteria for soil organisms in the framework of classification of substances and PBT assessment

The environmental hazard-based classification systems for chemicals, i.e., the Globally Harmonized System (GHS) and the European Classification, Labelling and Packaging Regulation (CLP), are currently based only on aquatic toxicity criteria. There are no legal requirements nor criteria to use toxicity data for soil organisms in this context, although soil toxicity data are available from several chemical regulations (such as plant protection products, pharmaceuticals, biocides, or REACH chemicals). It was assumed that aquatic criteria are conservative enough to sufficiently protect and inform hazards to terrestrial and soil compartments in the CLP context. However, exposure, uptake, and toxicity of chemicals differ between terrestrial and aquatic standard test organisms and -systems.

The PROSOIL project built a comprehensive harmonized database of ecotoxicity data for various soil organisms from four international data repositories (for the first time in such a comprehensive manner) to derive toxicity thresholds for soil organisms separated into five criterion groups. Moreover, candidate chemicals, which would be classified according to soil-toxicity criteria derived within this project from the soil ecotoxicity database *prosoildat,* have been identified.

After extensive data processing steps for harmonization, standardization, and quality assurance, further analyses were conducted on five defined coherent criterion groups, based on a combination of, e.g., test species and exposure durations (i.e., plants acute, plants chronic, soil macro-organisms acute, soil macro-organisms chronic, soil microorganisms chronic).

After applying strict quality criteria to the data, 125.000 observations in the PROSOIL database were available for analysis, covering 3.700 substances and 2.700 test species.

Toxicity threshold values were derived using three different statistical methodologies (geometric mean/quantile/Null Hypothesis Significance Testing). The chemicals whose effect value fell below the respective thresholds were "very toxic", "toxic", and "harmful" to the terrestrial environment. Applying these threshold values to the PROSOIL database resulted in the lists of candidate substances for each of the five groups, i.e., substances classified into one of the three hazard classes.

Finally, the substances on the candidate lists were compared to the respective existing aquatic classification within the European CLP inventory. The main result of this analysis was that a significant portion of the candidate substances was classified as hazardous to soil organisms based upon the PROSOIL data, although the existing aquatic classification according to CLP did not cover it. Hence, the environmental hazards of chemicals seem to be underestimated based on the existing classification system for aquatic organisms.

Kurzzusammenfassung: Schutz der Bodenorganismen: Entwicklung von Toxizitätskriterien für Bodenorganismen im Rahmen der Einstufung von Stoffen sowie der PBT-Bewertung

Die gefahrenbasierten Klassifizierungssysteme für Chemikalien, wie das Globale Harmonisierte System (GHS) sowie die europäische Verordnung zur Einstufung, Kennzeichnung und Verpackung (CLP), basieren derzeit in Bezug auf Umweltgefahren nur auf Kriterien für die aquatische Toxizität. Es gibt keine rechtlichen Anforderungen oder Kriterien für die Verwendung von Toxizitätsdaten für Bodenorganismen in diesem Zusammenhang, obwohl Daten zur Bodentoxizität aus vielen Regulierungsbereichen (wie z.B. Pflanzenschutzmittel, Arzneimittel, Biozide oder REACH-Chemikalien) verfügbar sind. Es wurde bisher davon ausgegangen, dass die Kriterien für Wasserorganismen konservativ genug sind, um über die Gefahren für terrestrische (und Boden-) Kompartimente im Rahmen der CLP-Verordnung zu informieren und einen ausreichenden Schutz der dort lebenden Organismen zu gewährleisten. Exposition, Aufnahme und Toxizität von Chemikalien gegenüber terrestrischen und aquatischen Standardtestorganismen und -systemen ist jedoch sehr unterschiedlich.

Im Rahmen des PROSOIL-Projekts wurde eine umfassende harmonisierte Datenbank mit Ökotoxizitätsdaten für verschiedene Bodenorganismen aus vier internationalen Datenbanken erstellt (zum ersten Mal in einer derart umfassenden Form), um Toxizitätsschwellenwerte für terrestrische Organismen, klassifiziert in fünf Kriterien-Gruppen, abzuleiten. Basierend auf der Bodenökotoxizitätsdatenbank *prosoildat* wurden Boden-Toxizitätskriterien abgeleitet, betroffene Substanzen identifiziert und in Kandidatenlisten zusammengefasst.

Nach umfangreichen Datenverarbeitungsschritten zur Harmonisierung, Standardisierung und Qualitätssicherung erfolgten weitere Analysen für fünf definierte kohärente Kriterien-Gruppen, auf Grundlage von u.a. Testarten und Testexpositionsdauern (Pflanzen akut, Pflanzen chronisch, Bodenmakroorganismen akut, Bodenmakroorganismen chronisch, Bodenmikroorganismen chronisch).

Nach Anwendung strenger Qualitätskriterien auf die Daten standen 125.000 Datensätze in der PROSOIL-Datenbank für Analysen zur Verfügung, die 3.700 Substanzen und 2.700 Testarten abdecken.

Die Toxizitätsschwellenwerte wurden mit Hilfe von drei verschiedenen statistischen Methoden (geometrisches Mittel / Quantil / Nullhypothesen-Signifikanztest) ermittelt. Die Chemikalien, deren Effektwert unter die festgelegten Schwellenwerte fiel, wurden als sehr toxisch ("very toxic"), toxisch ("toxic") und gefährlich ("harmful") für die terrestrische Umwelt eingestuft. Mit der Anwendung dieser Schwellenwerte auf die Daten in der PROSOIL-Datenbank wurden Listen von Kandidatenstoffen für jede der fünf Gruppen erstellt.

Die Stoffe der Kandidatenlisten wurden abschließend mit den entsprechenden aquatischen Einstufungen verglichen, die im europäischen CLP-Inventar festgelegt sind. Das wichtigste Ergebnis dieser Analyse war, dass ein erheblicher Teil der untersuchten Stoffe eine Einstufung gemäß der angewendeten PROSOIL Kriterien erhielt, aber nicht von der bestehenden aquatischen Klassifizierung unter CLP abgedeckt war. Folglich werden die Umweltgefahren von Chemikalien auf der Grundlage des bestehenden Klassifizierungssystems für Wasserorganismen bei Weitem unterschätzt.

Table of content

List of figures

List of tables

List of annexes

List of abbreviations

AMG	German Medicines Act
API	Application Programming Interface
ASTM	American Society for Testing and Materials
CAS	Chemical Abstracts Service
CLH	Harmonized Classification and Labelling
CLP	Classification, Labelling and Packaging
CM	carcinogenic, mutagenic, reprotoxic
sdw	soil dry weight
EC	Effect Concentration
ECHA	European Chemicals Agency
EMA	European Medicines Agency
EPPO	European and Mediterranean Plant Protection Organization
ETOX	Information System Ecotoxicology and Environmental Quality Targets
EU	European Union
EUCLEF	EU Chemicals Legislation Finder
GBIF	Global Biodiversity Information Facility
GHS	Global Harmonized System
HC_5	Hazard Concentration for five percent of the Community
ICS	Information Systems on Chemical Safety
INRA	Institute National de la Recherche Agronomique
IQR	Inter Quartile Range
ISO	International Standardisation Organisation
IUCLID	International Uniform Chemical Information Database
IUPAC	International Union of Pure and Applied Chemistry
L(E)C	Lethal (Effect) Concentration
LC_x	Lethal Concentration causing x percent of effect
LOEC	lowest observed effect concentration
mg ai / kg sdw	milligram of active ingredient per kg soil dry weight
NOEC	No Effect Concentration
NZEPA	New Zeeland Environmental Protection Agency
OECD	Organisation for Economic Co-operation and Development
PBT	Persistent, Bioaccumulative and Toxic substances
POP	Persistent Organic Chemicals
PPP	Plant Protection Product
PROSOIL	Protection of Soil Organisms
PUG-REST	Power User Gateway to Representation State Transfer
RAC	Committee for Risk Assessment, Regulatory Acceptable Concentration
REACH	Registration, Evaluation, Authorisation and Restriction of Chemicals
RSC	Royal Society of Chemistry
sdw	soil dry weight
SSD	Species Sensitivity Distribution
SVHC	Substance of Very High Concern

US EPA	United States Environmental Protection Agency
UBA	German Environment Agency
XML	eXtensible Markup Language

Summary

Background and motivation

Background on historical and recent initiatives to develop toxicity criteria for terrestrial organisms in the context of classification and labeling and PBT assessment.

Environmental hazard assessments in CLP and PBT are based on aquatic toxicity criteria. Around 15 years ago, **initiatives** and proposals to establish terrestrial toxicity criteria for classification and labeling at the EU and UN levels (European Commission, 2001) led to a proposal at the UN level in 2006 (Committee of Experts on the Transport of Dangerous Goods and the Globally Harmonized System of Classification and Labelling of Chemicals). These initiatives, however, were stopped in 2008 due to political reasons, cost-benefit considerations, and a lack of data. In 2020, the EU Commission announced in its "Chemical Strategy for Sustainability" to check the feasibility of including terrestrial toxicity criteria in the CLP regulation.

The **PROSOIL project** was built upon literature research of available proposals and analyses and aimed for the compilation of a comprehensive database on soil ecotoxicity data, the identification of toxicity ranges for different soil organism groups to propose toxicity criteria for use in the context of the classification and labeling regulation (CLP) and PBT assessments. The project focused on terrestrial organism groups exposed to toxic substances via soil, i.e., plants, invertebrates (e.g., oligochaetes, collembolans, mites), and microorganisms. Above-ground organisms such as bees, birds, or mammals were not considered. A comprehensive set of soil ecotoxicity data for chemicals across different chemical products' legislation (REACH chemicals, plant protection products, biocides, and pharmaceuticals) was considered, with varying data availability, quality, and coverage due to different data requirements within the respective regulations.

The PROSOIL database

Overview of the PROSOIL database: data quality and availability, groups defined for analyses

For the compilation of the PROSOIL database, a **comprehensive set of ecotoxicity data** (around 125.000 entries) for various organisms (plants, in-soil microorganisms, in-soil macroorganisms) were retrieved from four different regulatory repositories (data entries before harmonization, quality checks, and duplicate removal: UBA ICS [38.000], ECHA IUCLID [43.000], UBA ETOX [3.000] and US EPA's ECOTOX [131.000]). Data were harmonized concerning unified species identifiers, endpoint nomenclature, and test concentration units. Non-plausible outliers were discarded to ensure reliable distribution of ecotoxicity values. About 50 % of the data entries were not useable due to missing entries regarding unique identifiers (CAS numbers), test species, ecotoxicological values, or duplication. Most data entries were retrieved for plants and lumbricids, with plant protection products (active substances) reported as the dominant substance class. A majority of data were retrieved from the US EPA ECOTOX, particularly for plants and chronic data for soil macroorganisms.

The dataset analyzed 15 years ago focused on a dataset of around 600 entries for industrial chemicals. For the comprehensive data set compiled in the present project, the **challenges regarding data quality** are an essential obstacle to getting broad support for a new criterion. Large parts of the data retrieved had not undergone validity checks by regulatory authorities, such as non-standard studies from peer-reviewed scientific literature (e.g., from US EPA ECOTOX) or studies submitted by companies (ECHA IUCLID). On the contrary, some pesticide data (e.g., from UBA ETOX & ICS) comprised more tests and were checked in more detail.

Whether a missing regulatory review is a decisive quality criterion remains for discussion elsewhere. As some source databases contained confidential data, e.g., on active substances for plant protection products, biocides, or pharmaceuticals, it was not made publicly available.

Estimating plausible toxicity ranges

Statistical approaches to estimate threshold values, toxicity ranges for different soil criterion groups, and plausible soil toxicity criteria

Critical aspects of exploratory data analyses were presented to understand the data set and **distributions of ecotoxicity values** for different homogenous groups of soil organisms and decisions on the selection of groups. The distributions of ecotoxicity values were considered like pool data from various data sources and taxa. For example, the invertebrate taxa *Collembola, Lumbricidae, Enchytraeidae, Nematoda, Isopoda, Myriapoda, and Acari* were grouped as organism groups of "in-soil macroorganisms". Thereby, it was differentiated between acute and chronic data. Hence, five criterion groups were subjected to further analyses, threshold derivation, and definition of hazard categories: plants acute and plants chronic, macroorganisms acute and macroorganisms chronic, and microorganisms chronic. Acute data for macroorganisms was available primarily for earthworms, with *Eisenia fetida* being the most frequent test species but often not the most sensitive one. The sensitivity patterns were shown by various species sensitivity distributions.

Different approaches to data usage were compared to derive **ecotoxicity thresholds**, i.e., the "geomean", the "quantile", and the "null-hypothesis significance testing - NHST" approach.

The approaches reflect different basic definitions of a hazardous substance.

The **geomean approach** assumes that a substance must be assigned as hazardous if it is *more toxic than an average substance of the statistical population of all substances*. The **quantile approach** assumes that a substance must be assigned as hazardous if it is *more toxic than a distinct proportion of the total ranked population of substances*. The **NHST approach** assumes that a substance must be assigned as hazardous if it is *significantly more toxic than the mean population of substances*. Based on the derived thresholds for each approach, the distinct categories *harmful*, *toxic,* and *very toxic* were built (Geomean: toxic = geomean, very toxic = geomean/10, harmful = geomean*10; Null Hypothesis Significance Test: toxic = significance threshold < 5% error probability, very toxic = significance threshold/2, harmful= significance threshold *2; quantile: toxic = lower 10th quantile, very toxic = quantile/2, harmful = quantile*2). The number of substances falling below the threshold varied between the different statistical approaches. Statistical stability simulations (bootstrapping procedures) were conducted to check if the different approaches led to different conclusions for the case that fewer data would be available (down-sampling).

For five defined criterion groups, the toxicity ranges, and thresholds for the hazard classes "very toxic", "toxic," and "harmful", as well as the results of the geomean, the quantile, and the significance testing approaches were compared. Substances were then classified in **"candidate lists"** for each organism group (separately for each testing approach), listing those substances that would be classified as falling below the respective toxicity threshold. The differences between the five criterion groups were evaluated concerning the identities of candidate substances. Due to its overall lowest threshold, the geomean approach would lead to the highest number of classified substances. Often, the most sensitive species was *Eisenia fetida.* The highest percentage of substances classified for all defined criterion groups were regulated as plant protection products, followed by REACH chemicals or those taken from literature data (which could not be attributed to a regulatory area). At the same time, pharmaceuticals and biocides were far less represented (also, less data was available in the database).

In comparison to the terrestrial toxicity criteria of the previous proposal at the UN level in 2006, the thresholds derived by the PROSOIL project using the geomean approach were in similar orders of magnitude for most of the data groups (factor 2-5 lower than UN proposal), except for plants (factor 10-50 lower than UN proposal). Estimated thresholds (very toxic, toxic, and harmful) for the geomean approach were given as “mg ai/kg sdw” units.

- plant acute (EC_{50}) 0.2 - 1.6 - 16.2 vs UN 2006: 10 – 100 - 1000
- plant chronic (NOEC) 0.1 - 0.8 - 8.1 vs UN 2006: 1 – 10 - 100
- soil-macro acute (EC_{50}) 5.5 - 55.3 - 552.7 vs UN 2006: 10 – 100 - 1000
- soil-macro chronic (NOEC) 2.6 - 25.9 - 258.9 vs UN 2006: 1 – 10 - 100
- microorganisms (EC_{50}) 0.2 - 2.0 - 19.7 vs UN 2006: 1 – 10 - 100

Stakeholder involvement

An international workshop was organized as part of the validation and verification of the project´s approaches that aimed at putting the issue of integrating soil toxicity criteria into CLP and PBT assessments on the agenda again and at discussing options and data needed to derive toxicity criteria for soil organisms based on a new comprehensive database, as well as new analyses and proposals elaborated during the PROSOIL project. More than fifty experts from regulatory bodies and scientific institutions discussed the topic and added valuable ideas to the project and beyond.

In preparation for the workshop, participants were asked to fill in a *short questionnaire* on their **motivation, personal expertise, and viewpoints**, i.e., data quality, toxicity thresholds, and the need for additional protection for soil organisms. The results showed that most participants were experienced in environmental assessments focusing on the soil compartment. While the current protection level for in-soil organisms was better under the pesticide regulation than under CLP and REACH, most experts agreed that the inclusion of terrestrial criteria for CLP and PBT would be beneficial, and a future “soil framework directive” would be desirable. It was agreed that the derivation of threshold values should be as transparent as possible, that diverse data sources could be considered suitable for an overall generic assessment, but that data quality was crucial. Chronic and more complex ecotoxicological test data are still comparatively scarce.

During the plenary discussion, the **current plans of the EU Commission** to consider the inclusion of terrestrial criteria into the current CLP and UN GHS classification (e.g., harmful to the aquatic **and** terrestrial environment) were presented. A first proposal is envisaged for 2022, followed by discussions and agreements at the UN GHS level. Experts involved in the previous initiatives at the EU and UN levels to establish soil toxicity criteria for CLP/GHS described the main challenges mentioned: data availability and quality, and the importance of demonstrating the benefits were crucial. It was agreed that the workshop's conclusions constitute an essential trigger for further discussions and could support the development of proposals and decision-making processes.

It was also discussed how data quality could be best defined. Considering the enormous dataset, the reliability of the individual test results was not checked in detail in the PROSOIL project, but statistical methods were applied instead. It was assumed that the available data had undergone an assessment before uptake in the database, although it is known that data quality varies between the different databases. For the PROSOIL database, efforts were undertaken to harmonize the units of ecotoxicity endpoints to soil concentrations in mg active ingredient/kg dry weight of soil whenever possible. Moreover, it must be considered that data was compiled from diverse data sources and different test protocols within each of the four individual databases.

Plenary discussions were centered on a **better understanding and transparency of the reliability and comparability of the data** as an essential basis for any further analysis and the derivation of thresholds and toxicity ranges. During the plenary discussions, it was proposed to flag the **reliability of data entries** and information on the experimental conditions in future versions of the database and discuss biases, to strengthen confidence in the data basis. The need for improved availability and harmonization of data internationally was pointed out. A correction in the sense of normalization (e.g., by temperature) was also proposed. Moreover, it was discussed whether the analysis of a more minor but well-known and checked sub-set of data could be beneficial or whether statistical approaches are sufficient to understand data reliability. Further discussions are also needed concerning the **approaches used for deriving thresholds and hazard categories**. While the workshop presentations exemplified the results of the geomean approach, some participants pleaded for a probabilistic approach.

Finally, it was considered essential to gather **arguments to demonstrate the potential benefits** of introducing terrestrial toxicity criteria. The comparison with the existing aquatic hazard classification was emphasized as a measure, e.g., the possible number of substances that were not yet classified. A comparison of a similar analysis with the same approach used for deriving the aquatic thresholds was proposed. The classification system was meant to discriminate and highlight hazardous substances to the environment. However, not only should the sole number of additional classifications be taken as an argument, but also that terrestrial criteria are relevant to ensure the protection of soil organisms. The main task of CLP as a hazard communication tool (e.g., for transport, occupational health, substance authorizations, and assessments) was highlighted, together with the consequences of new terrestrial criteria for the consideration of the soil compartment in the future legislation.

Consequences, relevance, and benefits
Possible toxicity criteria and their "consequences": relevance and benefits for a better consideration of terrestrial organisms

To evaluate the **potential benefits and the relevance of toxicity criteria for soil organisms**, the potential terrestrial classifications for candidate substances were compared with the existing aquatic classifications (H-phrases from the CLP inventory on self-notified or harmonized classifications). It was analyzed whether a candidate substance classified as harmful, toxic, or very toxic to soil organisms was already classified in the same or a more critical aquatic hazard category. The comparison was conducted for each criterion group, respectively. Overall, these analyses highlighted the benefits of additional terrestrial toxicity criteria as a minority. However, many substances relevant to the terrestrial environment were still not yet classified by the existing system based on ecotoxicity data of aquatic organisms.

Moreover, it was also shown that only a low percentage of candidate substances were already identified as POPs, PBTs, vPvBs, and SVHCs.

Conclusions

Protection level of current CLP criteria and availability of soil toxicity data

The main driver of the present research was the **protection of the populations and the biodiversity of soil organisms** from adverse effects that arose from chemical pressures in the environment. The project's initiation was based on past activities on how terrestrial and soil criteria should be integrated into hazard characterization and provisions for classification and labeling or PBT assessments. The protectiveness of the current criteria, which were only based on aquatic data, was analyzed. It had to be concluded that specific soil-based toxicity criteria urgently needed to be discussed in a broader regulatory and scientific context. It should probably be implemented in European and international law wherever chemical legislation utilizes hazard-based criteria. Knowledge of toxicity thresholds of chemical substances for the environment may benefit prioritizations for further regulation.

It was shown that across the three statistical approaches applied for the derivation of toxicity thresholds, 10 – 30 % of all substances within the *prosoildat* database across the five criterion groups were not classified by the aquatic criteria within CLP. The absolute proportion of classified substances depended on the concrete threshold values. However, none of the PROSOIL approaches led to a complete match with the existing classification. The differences did not exclusively occur at the sensitive end of the toxicity distribution at low values, but the sensitivity patterns of soil organisms' toxicity endpoints turned out to be different from the aquatic test organisms. This finding determines that no simple shift of the existing (aquatic) threshold or the new (soil) threshold would lead to complete coverage of the two classification approaches, but a separate or integrated assessment of soil organisms was considered necessary. Furthermore, the lists of candidate substances in the highest hazard level of the developed soil toxicity criteria showed a low level of coincidence with existing exclusion criteria of the PBT, SVHC, or POP lists.

The toxicity thresholds suggested, derived by the three approaches, geomean, nhst, and quantile, were approximately in agreement (except for plants) with the thresholds proposed by historical approaches concerning their absolute values and, accordingly, their protectiveness. However, the documents, sources from discussions at the EU and UN level, and the data basis leading to the UN proposal in 2006 were not publicly available and, therefore, difficult to follow and compare. The only source with a clearly stated data basis (Renaud et al., 2004) came for the criterion group "soil macro chronic" to a similar estimation of thresholds as in the present study.

The relatively high number of substances that were not yet classified but would need to be labeled as hazardous to soil organisms to better inform on their hazards, resulting in a **low protection level of the current CLP criteria for aquatic organisms,** was a potentially alarming result of the research project (though depending on the selected thresholds). If confirmed, this would underpin the necessity of additional specific criteria for the soil compartment and should require attention and action from all stakeholder groups.

Terrestrial toxicity criteria were also seen as beneficial in PBT assessments. The T-criterion for PBT assessment could be the very toxic class or extended to a lowered threshold using a specific conversion factor. Within PBT assessments, it would allow more data to be used to fulfill the T-criterium, leading to a higher number of substances identified as PBT. And finally, the terrestrial compartment and associated data would be equally considered, which is currently primarily ignored under CLP and PBT assessments. Going beyond the CLP and PBT context, information on hazards to the soil compartment from the prospective assessments might also be necessary for the media-oriented frameworks to identify potential priority substances for monitoring or other measures.

The **availability of soil-related toxicity data** was much higher than expected from scientists involved in the project and the broader group of experts involved in the workshop discussions. The development of terrestrial cut-off criteria, e.g., by Carbonnell et al. in the year 1997, comprised approximately 1.000 chemical substances across all terrestrial assessment areas, including plants, birds & mammals, and earthworms. The *prosoildat* database held more than 3.700 different substances that covered a representative range of "intrinsic" physicochemical properties and the regulatory areas relevant to the CL of chemicals. Overall, the **PROSOIL database** was a significant step toward developing reliable, generalizable soil toxicity criteria offering a dataset of comprehensibility, with most data available for plant protection products and REACH chemicals. All the data analysis steps were documented in detail by making the algorithms available via human-readable programming scripts.

Nonetheless, the database came with some characteristics and limitations that should be considered when interpreting the results. One of the most extensive data packages of the PROSOIL database originated from the US EPA ECOTOX system, which collected data from peer-reviewed literature (see sections 3.2, 4.3). This data supposedly lacked the formal validity and reproducibility criteria applied in a regulatory context (i.e., checked by authorities). On the other hand, the data was checked during the peer-review process by at least two or more independent experts in the field for the scientific excellence and the reliability of the (experimental and statistical) methodology. The thorough, in-deep review process and its broader scope in test species and chemicals made the open-literature data particularly useful for scientific exercises.

The most extensive part of the data was linked to the regulation of plant protection products which undoubtedly influenced the evaluation of the complete data set. The data requirements for the regulation of plant protection products included more data sets on organisms related to the compartment soil than, e.g., chemicals regulated within REACH or biocides. Therefore, the toxicity thresholds' derivation of terrestrial organisms within the evaluated data set might be influenced by organic chemicals used as plant protection products. However, this bias did not hamper the protectiveness of the overall evaluation of soil toxicity criteria.

The **usability of data** from various origins and a long history of developing soil ecotoxicological test protocols further hampers the analysis due to their enormous heterogeneity. However, no general semantics and guidelines providing ontologies and vocabularies of soil ecotoxicological data were ready-for-use. It was a vast and time-consuming step to make the four data sources interoperable by pre-defining common notations of endpoints, species, and guidelines. The harmonization processes assured that the PROSOIL project could impact the long term beyond its duration.

After the involvement of an experienced group of experts at the conducted international workshop, including those that were engaged in the first attempts to establish soil criteria, it was agreed that the following steps after the finalization of the PROSOIL project would enhance the transparency and reliability of the existing data rather than the integration of additional data. To improve the **reliability of the data**, future projects should consider information on the reliability of origin of the data point, the guideline conformity, and usability for meta-analyses. Statistical tools could be used for quality checks in general. The use of smaller, well-described and reliable datasets would be considered beneficial for some types of analyses (i.e., when it comes to concrete regulatory measures).

For the defined **criterion groups**, it was concluded that the groups "plants", "macroorganisms," and "microorganisms" would be appropriate for use in sub-categories to derive terrestrial toxicity criteria. Other groups that might be considered were above-ground organisms, such as bees, mammals, or birds, which were not analyzed in the PROSOIL project due to the focus on organisms exposed via soil.

The PROSOIL project aimed to contribute to the ongoing discussions at the EU and UN levels to develop terrestrial toxicity criteria. The outcomes fit well into the ongoing activities within the **EU policy-making** processes. The European Commission is currently developing strategies for integrating soil toxicity criteria into the CLP and GHS regulations. Most stakeholders at the international workshop supported a future introduction of terrestrial toxicity criteria, but specific details, such as the analysis of well-defined data sets or how to consider above-ground organisms (e.g., bees, mammals), need further discussion.

Also, the approaches to derive the **toxicity thresholds** and the use of the derived criteria need further discussion, particularly in comparison with the analyses and proposals of the previous initiatives at the UN level.

Although the PROSOIL project focused mainly on the derivation of thresholds directly translated into hazard classes for CLP, terrestrial criteria were also seen as beneficial in PBT assessments.

Zusammenfassung

Hintergrund und Motivation

Hintergrundinformationen zu den bisherigen Initiativen zur Entwicklung von Toxizitätskriterien für terrestrische Organismen im Zusammenhang mit Einstufung und Kennzeichnung (CLP) sowie der PBT-Bewertung von Chemikalien.

Derzeit basieren Ökotoxizitätsbewertungen im Rahmen von CLP und PBT ausschließlich auf aquatischen Toxizitätskriterien. Vor etwa 15 Jahren gab es mehrere Initiativen und Vorschläge zur Festlegung terrestrischer Toxizitätskriterien für die Einstufung und Kennzeichnung von Chemikalien auf EU- und UN-Ebene (European Commission 2001). Diese führten zu einem Vorschlag auf UN-Ebene im Jahr 2006 (Committee of Experts on the Transport of Dangerous Goods and the Globally Harmonized System of Classification and Labelling of Chemicals), welcher jedoch im Jahr 2008 aus politischen Gründen, aus Kosten-Nutzen-Erwägungen und wegen einer mangelhaften Datengrundlage gestoppt wurde. Im Jahr 2020 kündigte die EU-Kommission in ihrer „Chemikalienstrategie für Nachhaltigkeit" an, die Machbarkeit einer Aufnahme von terrestrischen Toxizitätskriterien in die CLP-Verordnung zu prüfen.

Das PROSOIL-Projekt hatte das Ziel, aufbauend auf einer Literaturrecherche bereits veröffentlichter Studien und Analysen, eine umfangreiche Datenbank von Bodenökotoxizitätsdaten zusammenzustellen, die eine Identifizierung von Toxizitätsbereichen für verschiedene Gruppen von Bodenorganismen und die Ableitung von Toxizitätskriterien für die Verwendung im Zusammenhang mit der Einstufung und Kennzeichnung von Chemikalien (CLP) und PBT-Bewertungen ermöglicht. Das Projekt konzentrierte sich auf terrestrische Organismengruppen, die über den Boden toxischen Substanzen ausgesetzt sind, d. h. Pflanzen, wirbellose Tiere (z. B. Oligochaeten, Collembolen, Milben) und Mikroorganismen. Andere terrestrische Organismen, für die keine Exposition über den Boden zu erwarten ist, z.B. Bienen, Vögel oder Säugetiere wurden zunächst nicht berücksichtigt. Insgesamt wurde eine umfassende Sammlung von Bodenökotoxizitätsdaten für Chemikalien aus verschiedenen Regulierungsbereichen (REACH, Pflanzenschutzmittel, Biozide und Arzneimittel) berücksichtigt, wobei die Datenverfügbarkeit, -qualität und -abdeckung aufgrund unterschiedlicher regulatorischer Datenanforderungen variierte.

Die PROSOIL-Datenbank

Überblick über die PROSOIL-Datenbank: Inhalt, Datenqualität und Datenverfügbarkeit

Für die Erstellung der PROSOIL-Datenbank wurde ein umfassender Satz von Ökotoxizitätsdaten (ca. 125.000 Einträge) für verschiedene Organismen (Pflanzen, Mikroorganismen im Boden, Makroorganismen im Boden) aus vier verschiedenen regulatorischen Datenrepositorien (Dateneinträge vor Harmonisierung, Qualitätsprüfung und Entfernung von Duplikaten: UBA ICS [38.000], ECHA IUCLID [43.000], UBA ETOX [3.000] und ECOTOX der US-EPA ECOTOX [131.000]) zusammengestellt. Die Daten wurden in Bezug auf einheitliche Einheiten, z.B. Artenbezeichnungen, Endpunktnomenklatur und Testkonzentrationseinheiten, harmonisiert. Nicht plausible Ausreißer wurden aussortiert, um eine zuverlässige Verteilung der Ökotoxizitätswerte zu erhalten. Etwa 50 % der Daten waren aufgrund fehlender Einträge für eindeutige Identifier (z.B. CAS-Nummern), Testarten oder Effektwerte oder aber aufgrund von duplizierten Einträgen nicht nutzbar. Die meisten Dateneinträge wurden für Pflanzen und Lumbriciden abgerufen, wobei Pflanzenschutzmittel (Wirkstoffe) die am häufigsten vorkommende Stoffklasse darstellen. Die größte Anzahl von Datensätzen wurde aus der US-EPA ECOTOX Datenbank abgerufen; hier insbesondere für Pflanzen und chronische Daten für

Bodenmakroorganismen. Die vor 15 Jahren in früheren Studien analysierten Datensätze umfassten lediglich maximal 600 verschiedene Chemikalien.

Eine wichtige Problematik der diesem Projekt zugrunde liegende Daten liegt in den Herausforderungen hinsichtlich der Qualität. Dies ist insbesondere wichtig um eine breite Unterstützung für ein neues (Boden-) Toxizitätskriterium zu erhalten. Ein großer Teil der abgerufenen Daten wurde nicht von Regulierungsbehörden auf ihre Vertrauenswürdigkeit hin geprüft. Häufig liegen nicht standardisierte Studien aus wissenschaftlicher Literatur (peer-reviewed) oder von Unternehmen selbst vorgelegte Studien (ECHA IUCLID) vor. Im Gegensatz dazu wurden einige Pestiziddaten (UBA ETOX) behördlich geprüft. Ob eine fehlende regulatorische/behördliche Begutachtung ein ausschlaggebendes Qualitätsmerkmal ist, bleibt anderweitig zu diskutieren. Eine vollständige Veröffentlichung der PROSOIL Datenbank ist nicht vorgesehen, da die UBA ICS Datenbank vertrauliche Daten enthält, z. B. über Wirkstoffe für Pflanzenschutzmittel, Biozide oder Arzneimittel, und dementsprechend nicht öffentlich zugänglich gemacht werden darf.

Prognose von Toxizitätsbereichen

Statistische Ansätze des PROSOIL-Projekts zur Prognose von Schwellenwerten, Bodentoxizitätskriterien und Toxizitätsbereichen für verschiedene Kriterien-Gruppen

Eine explorative Datenanalyse wurde auf die PROSOIL-Datenbank angewendet, um ein Verständnis des Datensatzes und der Verteilungen der Ökotoxizitätsdaten für verschiedene homogene Gruppen von Bodenorganismen zu gewinnen und darauf basierend Entscheidungen über die Auswahl der Gruppen treffen zu können. Die homogene Verteilung der Ökotoxizitätsdaten erlaubte es, Daten aus verschiedenen Quellen und für unterschiedliche Organismen zusammenzufassen. So wurden beispielsweise wirbellose Taxa *Collembola, Lumbricidae, Enchytraeidae, Nematoda, Isopoda, Myriapoda* und *Acari* als Bodenmakroorganismen zusammengefasst. Dabei wurde zwischen akuten und chronischen Daten unterschieden. Insgesamt wurden so fünf Datengruppen (im folgenden Kriterien-Gruppen) für weiterführende Analysen, der Ableitung von Schwellenwerten und der Definition von Gefahrenkategorien definiert: Pflanzen akut und chronisch, Bodenmakroorganismen akut und chronisch sowie Mikroorganismen chronisch. Akute Daten für Bodenmakroorganismen lagen für hauptsächlich für Regenwürmer vor, wobei *Eisenia fetida* die häufigste Testart darstellt. Während die UBA ETOX Datenbank die meisten Einträge für akute Daten enthielt, dominierten in der US-EPA ECOTOX Datenbank chronische Daten für Bodenmakroorganismen.

Zur Ableitung von Schwellenwerten wurden verschiedene statistische Ansätze verglichen, namentlich „Quantil-Ansatz", „Geomean-Ansatz " und „Null-Hypothese-Signifikanztest". Auf der Grundlage der für jeden Ansatz abgeleiteten Schwellenwerte wurden die Kategorien „harmful“, „toxic“ und „very toxic“ gebildet (Geomean: toxic = geomean, very toxic= geomean/10, harmful=geomean*10; Null-Hypothese-Signifikanztest: toxic = Signifikanzschwelle < 5% Fehlerwahrscheinlichkeit, very toxic = Signifikanzschwelle /2, harmful= Signifikanzschwelle *2; Quantil: toxic = unteres 10. Quantil, very toxic = Quantil/2, harmful = Quantil*2). Die Anzahl der Stoffe, die unter die so definierten Schwellenwerte fallen, variiert zwischen den verschiedenen statistischen Ansätzen. Weiterhin wurden Simulationen zur statistischen Stabilität (Bootstrapping-Verfahren) durchgeführt, um zu prüfen, ob die verschiedenen Ansätze zu unterschiedlichen Schlussfolgerungen führen, wenn weniger Daten zur Verfügung stehen (Downsampling). Dies war nicht der Fall.

Für fünf Organismengruppen wurden die Toxizitätsbereiche und Schwellenwerte für die Gefahrenklassen "very toxic", "toxic" und "harmful" mit den Ergebnissen der Geomean-, Quantil- und Signifikanztest-Ansätze verglichen. Die Stoffe wurden dann für jede Organismengruppe in

"Kandidatenlisten" zusammengestellt, d.h. es wurden diejenigen Stoffe aufgelistet, die unterhalb der jeweiligen Toxizitätsschwelle eingestuft werden würden. Die Unterschiede zwischen den Kriterien-Gruppen wurden hinsichtlich der Kandidatenstoffe bewertet. Der Geomean-Ansatz führte zur höchsten Anzahl klassifizierter Substanzen. Eine der empfindlichsten Arten war häufig *Eisenia fetida*, da sie in den letzten Jahrzehnten als Standardtestart dominierte. Der höchste Prozentsatz an Stoffen, die für alle definierten Kriterien-Gruppen eingestuft wurden, waren Pflanzenschutzmittel, gefolgt von REACH Daten und Daten ohne spezifizierten Regulierungsbereich, während Pharmazeutika und Biozide weit weniger vertreten waren.

Im Vergleich zu den Toxizitätskriterien des früheren Vorschlags auf UN-Ebene aus dem Jahr 2006 lagen die im Rahmen des PROSOIL-Projekts unter Verwendung des Geomean-Ansatzes abgeleiteten Schwellenwerte für die meisten Datengruppen, mit Ausnahme von Pflanzen, in ähnlichen Größenordnungen. Geschätzte Schwellenwerte (very toxic - toxic - harmful) für den Geomean-Ansatz, angegeben in der Einheit "mg ai/kg sdw":

- Pflanze akut EC_{50} 0.2 - 1.6 - 16.2 vs UN 2006: 10 – 100 - 1000
- Pflanze chronisch NOEC 0.1 - 0.8 - 8.1 vs UN 2006: 1 – 10 - 100
- Bodenmakroorganismen akut EC_{50} 5.5 - 55.3 - 552.7 vs UN 2006: 10 – 100 - 1000
- Bodenmakroorganism. chronisch NOEC 2.6 - 25.9 - 258.9 vs UN 2006: 1 – 10 - 100
- Mikroorganismen chronisch EC_{50} 0.2 - 2.0 - 19.7 vs UN 2006: 1 – 10 - 100

Workshop und Einbeziehung von Interessenvertretern

Im Rahmen der Validierung und Verifizierung der Projektansätze wurde ein Workshop organisiert, der darauf abzielte, die Frage der Integration von Bodentoxizitätskriterien in CLP- und PBT-Bewertungen erneut auf die regulatorische Tagesordnung zu setzen und die Optionen und den Datenbedarf für die Ableitung von Toxizitätskriterien für Bodenorganismen auf der Grundlage einer neuen umfassenden Datenbank des PROSOIL-Projekts zu diskutieren. Mehr als fünfzig Experten auf diesem Gebiet, primär von Regulierungsbehörden und wissenschaftlichen Einrichtungen, waren involviert.

Zur Vorbereitung des Workshops wurden die Teilnehmer gebeten, einen kurzen Fragebogen zu ihrer Motivation, ihrem persönlichen Fachwissen und ihren Ansichten auszufüllen, so z. B. zur Datenqualität, zu Toxizitätsschwellenwerten und zur Notwendigkeit eines zusätzlichen Schutzes für Bodenorganismen. Die Ergebnisse zeigten, dass die meisten Teilnehmer Erfahrung mit Umweltbewertungen mit Schwerpunkt auf dem Kompartiment Boden hatten. Zwar wurde das derzeitige Schutzniveau für Bodenorganismen im Rahmen der Pestizidverordnung als besser empfunden als im Rahmen von CLP und REACH, doch waren sich die meisten Experten einig, dass die Einbeziehung terrestrischer Kriterien für CLP und PBT von Vorteil wäre, und dass ggfs. eine künftige "Bodenrahmenrichtlinie" wünschenswert wäre. Man war sich einig, dass die Ableitung von Schwellenwerten so transparent wie möglich erfolgen sollte, dass die unterschiedlichen Datenquellen als geeignet für eine generische Gesamtbewertung angesehen werden könnten und dass die Verfügbarkeit von Daten generell als entscheidend angesehen wird. Chronische und komplexere ökotoxikologische Testdaten sind noch immer vergleichsweise schlecht verfügbar.

In der Plenardiskussion wurden die aktuellen Pläne der EU-Kommission, die Aufnahme von terrestrischen Kriterien in die aktuelle CLP- und UN-GHS-Einstufung zu prüfen (z.B. schädlich für die aquatische und terrestrische Umwelt), präsentiert und für das Jahr 2022 in Aussicht gestellt, gefolgt von Diskussionen und Vereinbarungen auf UN-GHS-Ebene. Auch die laufenden

Aktivitäten zur Aufnahme von CLP-Kriterien für PBT und PMT wurden erwähnt. Weitere Experten, welche an den früheren Initiativen auf EU- und UN-Ebene zur Festlegung von Bodentoxizitätskriterien für CLP/GHS beteiligt waren, beschrieben die aus ihrer Sicht wichtigsten Herausforderungen: Die Verfügbarkeit und Qualität von Daten sowie die Bedeutung der Relevanz/Nutzen-Analyse waren damals von entscheidender Bedeutung. Es wurde festgehalten, dass die Schlussfolgerungen des Workshops einen wichtigen Anstoß für weitere Diskussionen auf EU/UN-Ebene darstellen und die Entwicklung unterstützen können.

Weiterhin wurde diskutiert, wie die Datenqualität am besten definiert werden kann. Angesichts der großen Datenmenge der PROSOIL-Datenbank wurde die Zuverlässigkeit der einzelnen Datensätze nicht einzeln überprüft, jedoch statistische Methoden angewandt, um die Qualität zu überprüfen. Die Diskussionen im Plenum konzentrierten sich auf ein besseres Verständnis und die transparente Darstellung von Zuverlässigkeit und Vergleichbarkeit der Daten als wichtige Grundlage für alle nachfolgenden Analysen, die Ableitung von Schwellenwerten und der Bestimmung von Toxizitätsbereichen. Für die PROSOIL-Datenbank wurden Anstrengungen unternommen, verschiedene Einheiten, z.B. der ökotoxikologischen Endpunkte, so weit wie möglich zu harmonisieren. Eine solche Harmonisierung der Daten soll eine Zusammenstellung verschiedenen Datenquellen und unterschiedlicher Testprotokolle ermöglichen. Auch eine Normalisierung der Daten (z.B. nach Temperatur) wurde vorgeschlagen. Weiterhin wurde diskutiert, ob die Analyse einer kleineren, aber bekannten und geprüften Teilmenge von Daten von Vorteil sein könnte, oder ob weitere statistische Ansätze angewandt werden könnten, um die Zuverlässigkeit der Daten zu verbessern oder zumindest besser zu verstehen. Weiterer Diskussionsbedarf besteht auch hinsichtlich der statistischen Ansätze zur Ableitung von Schwellenwerten und Gefahrenkategorien. Während die Ergebnisse des „Geomean“-Ansatzes in den Workshop-Präsentationen veranschaulicht wurden, plädierten einige Teilnehmer für einen probabilistischen Ansatz.

Bei den Diskussionen im Plenum wurde vorgeschlagen, die Zuverlässigkeit der Dateneinträge und der Informationen über die Testbedingungen in künftigen Versionen der Datenbank zu kennzeichnen, um das Vertrauen in die Datengrundlage zu stärken. Es wurde generell auf die Notwendigkeit einer verbesserten Verfügbarkeit und Harmonisierung von Daten auf internationaler Ebene hingewiesen. Außerdem wurde es als wichtig erachtet, Argumente zu sammeln, um die potenziellen Vorteile einer Einführung von Kriterien für die Bodentoxizität aufzuzeigen. Der durchgeführte Vergleich der PROSOIL-Daten mit der bestehenden aquatischen Gefahrenklassifizierung innerhalb von CLP wurde als Argument hervorgehoben. Das Klassifizierungssystem sollte dazu dienen, gefährliche oder giftige Stoffe zu unterscheiden und hervorzuheben. Allerdings sollte nicht allein die Anzahl der zusätzlichen Einstufungen als Argument herangezogen werden, sondern auch, dass terrestrische Kriterien als solche relevant sind, um den Schutz von Bodenorganismen zu gewährleisten. Die Hauptaufgabe der CLP-Verordnung als Instrument der Gefahrenkommunikation (z. B. für den Transport, den Gesundheitsschutz am Arbeitsplatz, die Zulassung und Bewertung von Stoffen) wurde ebenso unterstrichen wie die Folgen der neuen terrestrischen Kriterien für die Berücksichtigung des Kompartiments Boden in künftigen Rechtsvorschriften.

Konsequenzen, Relevanz und Nutzen

Bodentoxizitätskriterien und ihre Konsequenzen: Relevanz und Nutzen für eine bessere Berücksichtigung von terrestrischen Organismen

Um den potenziellen Nutzen und die Relevanz von Toxizitätskriterien für Bodenorganismen aufzuzeigen, wurden die Einstufungen für Kandidatenstoffe des PROSOIL-Projekts mit den bestehenden aquatischen Einstufungen (H-Sätze aus dem ECHA CLP-Verzeichnis; selbst angemeldete oder harmonisierte Einstufungen) verglichen. Hier wurde analysiert, ob ein innerhalb der PROSOIL-Daten als „harmful", „toxic" oder „very toxic" eingestufter Kandidatenstoff bereits in dieselbe oder nächst-strengere aquatische Gefahrenkategorie eingestuft wurde. Dies wurde für jede Kriterien-Gruppe separat analysiert. Insgesamt machten diese Analysen den Nutzen zusätzlicher terrestrischer Toxizitätskriterien deutlich, da eine Minderheit jedoch trotzdem große Anzahl von Stoffen, die für die terrestrische Umwelt relevant sind, im bestehenden CLP-System, das auf Ökotoxizitätsdaten aquatischer Organismen beruht, nicht eingestuft wurde. Außerdem zeigte sich, dass nur ein geringer Prozentsatz der in Frage kommenden Stoffe bereits als POPs, PBTs, vPvBs und/oder SVHCs identifiziert wurde.

Schlussfolgerungen

Schutzniveau der bisherigen CLP-Kriterien und Verfügbarkeit von Toxizitätsdaten für Bodenorganismen

Primäres Ziel des vorliegenden Projekts war der Schutz der Populationen und der biologischen Vielfalt von Bodenorganismen vor chemischen Belastungen in der Umwelt. Das Vorhaben wurde aufgrund und aufbauend auf früheren Aktivitäten zur Entwicklung von terrestrischen und Bodentoxizitätskriterien für die Einbindung in CLP und PBT initiiert. Die Schutzwirkung der derzeitigen Toxizitätskriterien, die nur auf aquatischen Daten beruhen, wurde hierzu analysiert, was zur Schlussfolgerung führte, dass spezifische bodenbasierte Toxizitätskriterien dringend in einem breiteren regulatorischen und wissenschaftlichen Kontext diskutiert werden müssen. Eine Umsetzung in europäisches und internationales Recht wäre überall dort sinnvoll, wo die Chemikaliengesetzgebung auf Vorsichtsmaßnahmen oder risikobasierte Kriterien verweist und wo Schwellenwerte für die Freisetzung chemischer Stoffe in die Umwelt erforderlich sind.

Das Projekt konnte zeigen, dass bei allen drei gewählten Ansätzen, die für die Ableitung von Toxizitätsschwellenwerten verwendet wurden, 10 bis 30 % aller Stoffe in der *prosoildat*-Datenbank über die fünf Kriterien-Gruppen hinweg nicht von den aquatischen Kriterien der CLP-Verordnung abgedeckt sind. Der absolute Anteil der eingestuften Stoffe hing zwar von den konkreten Schwellenwerten ab, keiner der Analyse-Ansätze führte jedoch zu einer vollständigen Übereinstimmung mit der bestehenden (aquatischen) CLP-Einstufung. Außerdem traten die Unterschiede nicht ausschließlich im sensitivsten Bereich der Toxizitätsverteilungen (bei niedrigen Werten) auf. Dieses Ergebnis zeigte, dass eine einfache Verschiebung des bestehenden (aquatischen) Schwellenwerts oder des neuen (Boden-) Schwellenwerts nicht zu einer vollständigen Abdeckung der beiden Klassifizierungsansätze führen würde, sondern dass eine separate oder integrierte Bewertung der Bodenorganismen für notwendig erachtet werden muss. Darüber hinaus zeigten die Listen der Kandidatenstoffe in der höchsten Gefahrenstufe („very toxic") der entwickelten Bodentoxizitätskriterien eine geringe Übereinstimmung mit den bestehenden Ausschlusskriterien der PBT-, SVHC- oder POP-Listen.

Die in Kapitel 6 vorgeschlagenen Toxizitätsschwellenwerte, die von den drei Ansätzen „geomean", „nhst" und „quantile" abgeleitet wurden, stimmten (mit Ausnahme von Pflanzen) hinsichtlich ihrer absoluten Werte und damit in ihrer Schutzwirkung annähernd mit den von früheren Ansätzen vorgeschlagenen Schwellenwerten überein. Allerdings waren die zugrundeliegenden Dokumente, Quellen und die Datengrundlagen, die zum UN-Vorschlag im Jahr 2006 führten, nicht öffentlich zugänglich und daher schwer nachzuvollziehen und zu vergleichen. Die einzige Quelle mit klar definierter Datengrundlage (Renaud et al. 2004) kam für die Kriteriums-Gruppe "Boden Makroorganismen chronisch" zu einer ähnlichen Abschätzung von Schwellenwerten wie in der vorliegenden Studie.

Die große Anzahl von Stoffen, die aktuell noch nicht aquatisch eingestuft wurden, aber als gefährlich für Bodenorganismen gekennzeichnet werden müssten, und somit zu einem niedrigeren Schutzniveau führt, war ein potenziell alarmierendes Ergebnis dieses Projekts (jedoch abhängig von den gewählten Schwellenwerten). Sollte es sich bestätigen, würde dies die Notwendigkeit zusätzlicher spezifischer Kriterien für das Kompartiment Boden untermauern und sollte die Aufmerksamkeit und das Handeln aller involvierter Interessengruppen fordern.

Terrestrische bzw. Bodentoxizitätskriterien wurden auch im Zusammenhang mit der PBT-Bewertung als nützlich erachtet. Das T-Kriterium für die PBT-Bewertung könnte entweder die Klasse „very toxic“ umfassen, oder unter Verwendung eines spezifischen Sicherheitsfaktors zu einem noch niedrigeren Schwellenwert erweitert werden. Im Rahmen der PBT-Bewertung könnte dies zu einer höheren Anzahl von Stoffen führen, die das T-Kriterium erfüllen und somit potenziell als PBT identifiziert werden würden. Schließlich würden das Kompartiment Boden und die damit verbundenen Daten in der PBT-Bewertung gleichermaßen berücksichtigt werden wie die Aquatik, was derzeit nicht der Fall ist. Über den CLP- und PBT-Kontext hinaus könnten Informationen über Gefahren für das Kompartiment Boden aus den prospektiven Bewertungen auch für andere regulatorische Rahmenwerke nützlich sein, um potenziell prioritäre Stoffe für die Überwachung oder andere Maßnahmen zu ermitteln.

Die Verfügbarkeit bodenbezogener Toxizitätsdaten war aus Sicht der an der Projektbearbeitung beteiligten Wissenschaftler und der breiteren Gruppe von Experten, die an den Workshop-Diskussionen teilnahmen, viel höher als erwartet. Die Entwicklung von terrestrischen Toxizitätskriterien, z.B. durch Carbonnell et al. im Jahr 1997, umfasste ca. 1.000 chemische Substanzen über alle terrestrische Bewertungsbereiche (Pflanzen, Vögel und Säugetiere sowie Regenwürmer) hinweg. Die PROSOIL-Datenbank enthielt mehr als 3.700 verschiedene Stoffe, die ein repräsentatives Spektrum an "intrinsischen" physikalisch-chemischen Eigenschaften und die für CLP relevanten Regulierungsbereiche abdeckten. Insgesamt wurde die PROSOIL-Datenbank als ein großer Schritt in Richtung der Entwicklung zuverlässiger, allgemeiner Kriterien für die Bodentoxizität angesehen, mit einem verständlichen Datensatz zu einem Großteil bestehend aus Daten für Pflanzenschutzmittel und REACH-Chemikalien.

Dennoch weist die PROSOIL-Datenbank einige Einschränkungen auf, die bei der Interpretation der Ergebnisse berücksichtigt werden sollten. Eines der umfangreichsten Datenpakete der PROSOIL-Datenbank stammt ursprünglich aus dem ECOTOX-System der US-EPA, welches hauptsächlich von Fachleuten überprüfte Literaturdaten enthält (peer-reviewed; siehe Abschnitte 3.3 und 4.3). Diesen Daten fehlen formale Gültigkeits- und Reproduzierbarkeitskriterien für eine Verwendung im regulatorischen Kontext (Überprüfung durch die Behörden). Andererseits wurden die Daten im Rahmen des Peer-Review-Verfahrens von mindestens zwei oder mehr unabhängigen Experten aus dem Fachgebiet auf ihre wissenschaftliche Qualität und die Zuverlässigkeit der (experimentellen und statistischen) Methodik überprüft. Der gründliche, eingehende Überprüfungsprozess und sein breiterer Anwendungsbereich in Bezug auf Testarten und Chemikalien machen die Daten aus der offenen Literatur deshalb besonders nützlich für wissenschaftliche Arbeiten.

Ein weiterer potenziell einschränkender Punkt ist, dass der umfangreichste Teil der PROSOIL-Daten sich auf die Regulierung von Pflanzenschutzmitteln bezog, was zweifellos die Bewertung des gesamten Datensatzes beeinflusste. Die Datenanforderungen für die Regulierung von Pflanzenschutzmitteln umfassten mehr Datensätze zu Organismen, die mit dem Kompartiment Boden in Verbindung stehen, als z. B. Chemikalien, die im Rahmen von REACH oder Bioziden reguliert werden. Deshalb könnte die Ableitung der Toxizitätsschwellenwerte für Bodenorganismen im ausgewerteten Datensatz durch organische Chemikalien, die als Pflanzenschutzmittel verwendet werden, beeinflusst sein. Diese Verzerrung in Richtung toxischerer Bereiche kann jedoch die Schutzwirkung der Gesamtbewertung durch Bodentoxizitätskriterien nicht beeinträchtigen.

Insgesamt enthält die finale PROSOIL-Datenbank einen großen ökotoxikologischen Datensatz für das Kompartiment Boden. Die statistischen Ansätze, die Datenverarbeitung sowie alle Schritte der Datenanalyse wurden detailliert dokumentiert, indem die Algorithmen über menschenlesbare Programmierskripte zur Verfügung gestellt wurden.

Die Verwendbarkeit dieser Daten unterschiedlicher Herkunft und unterschiedlicher ökotoxikologischer Testprotokolle führt jedoch zu einer Heterogenität, die die Analysen in gewissem Maße erschwert. Es waren zunächst zeitaufwändige Schritte notwendig, um die vier Quelldatenbanken kompatibel zu machen. Dies wurde durch eine Harmonisierung von Endpunkten, Einheiten und Artnamen gewährleistet, und somit sichergestellt, dass die PROSOIL-Datenbank auch in Zukunft weiterverwendet werden kann. Die Bemühungen und Ressourcen zur Implementierung von Bodentoxizitätskriterien in Gefahrenklassifizierungssysteme oder zu deren Verwendung bei der PBT-Bewertung und zur Priorisierung von Stoffen scheinen gerechtfertigt zu sein, um einen besseren Schutz von Bodenorganismen zu gewährleisten.

Während des Austauschs mit den Experten in einem internationalen Workshop, einschließlich derjenigen, die an den ersten Studien zu Bodentoxizitätskriterien beteiligt waren, wurde diskutiert, dass nach Abschluss des PROSOIL-Projekts eher die Verbesserung der Transparenz und die Zuverlässigkeit der vorhandenen Daten im Mittelpunkt stehen sollten, als die Integration zusätzlicher Daten voranzutreiben. Um die Glaubwürdigkeit der Daten zu verbessern, sollten künftige Projekte Informationen über die Zuverlässigkeit der Herkunft der Datenpunkte, die Richtlinienkonformität und die Verwendbarkeit für Meta-Analysen berücksichtigen. Statistische Instrumente könnten für allgemeine Qualitätskontrollen eingesetzt werden. Die Verwendung kleinerer, gut beschriebener und zuverlässiger Datensätze werden für einige spezielle Analysen als vorteilhaft angesehen (z.B. bei konkreten Regulierungsmaßnahmen).

Die definierten Kriterien-Gruppen des PROSOIL Projekts, also für akute und chronische ökotoxikologische Daten zu "Pflanzen", "Makroorganismen" und "Mikroorganismen", werden als für die Verwendung in Unterkategorien von terrestrischen Toxizitätskriterien geeignet angesehen. Im nächsten Schritt kämen als weitere Gruppen oberirdische Organismen wie Bienen, Säugetiere oder Vögel in Frage, die jedoch im Rahmen des PROSOIL-Projekts nicht analysiert wurden, da der Schwerpunkt auf Organismen lag, die über den Boden exponiert sind und im Boden leben.

Ziel des PROSOIL-Projekts war es, einen Beitrag zu den laufenden Diskussionen auf EU- und UN-Ebene über die Entwicklung von terrestrischen Toxizitätskriterien zu leisten. Die Ergebnisse des Projekts fügen sich gut in die aktuellen Aktivitäten im Rahmen der politischen Entscheidungsprozesse der EU ein. Die Europäische Kommission plant derzeit die Entwicklung von Strategien, wie Bodentoxizitätskriterien in die CLP- und GHS-Verordnungen integriert werden könnten. Eine künftige Einführung solcher Kriterien wurde auch von den meisten Teilnehmern des internationalen Workshops befürwortet, wobei die Details weiterer Diskussionen und Analysen bedürfen.

Auch die Ansätze zur Ableitung der Toxizitätsschwellenwerte und die Verwendung der abgeleiteten Kriterien müssen weiter diskutiert werden, insbesondere im Vergleich zu den Analysen und Vorschlägen der früheren Initiativen auf UN-Ebene. Obwohl sich das PROSOIL-Projekt hauptsächlich auf die Ableitung von Schwellenwerten konzentrierte, die direkt in Gefahrenklassen für die CLP-Verordnung umgesetzt werden, wurden auch terrestrische Kriterien als nützlich im Zusammenhang mit PBT-Bewertungen angesehen.

1 Background & introduction

1.1 Motivation and aims of this study

The classification and labeling of chemical substances as hazardous to the environment is focused on aquatic toxicity criteria. The criteria of the international Globally Harmonized System - GHS, as implemented in European law by the Classification, Labelling and Packaging – CLP regulation and the criteria for identifying persistent, bioaccumulative substances and toxic - PBT, do not include explicit or implicit provisions for soil toxicity. Inadequate provisions for terrestrial organisms were previously identified as a shortcoming regarding the protectiveness of the GHS/CLP regulations and PBT assessments for organisms other than fishes, aquatic plants, and aquatic invertebrates.

The deficits of the current system led to the main scientific question of this project: Are soil organisms adequately protected by hazard-based provisions derived from toxicity endpoints from aquatic test systems? It is questionable if there is a scientific rationale to assume that aquatic hazard criteria could protect all compartments of an ecosystem, be it aquatic (marine, limnic, sediment), terrestrial (above-ground), or soil (below-ground).

In the recent past, regulatory initiatives to strengthen the policies in soil protection had little success. In 2006, the European Commission established a thematic strategy for soil protection, an effort that should finally lead to the establishment of a European soil framework directive, equivalent to the water framework directive. The proposed directive was withdrawn in 2014. Previous efforts at the EU and international levels to integrate soil-specific classification and criteria in GHS/CLP frameworks or PBT assessments on the EU and UN levels have been examined and presented in chapter 1.3.

There are reasons to hypothesize that soil organisms are different in how a chemical substance could adversely affect them compared to aquatic organisms. The differences are due to different exposure and uptake routes and genetic variation between taxonomic groups of aquatic and terrestrial species. Taxonomic variability inevitably causes variation in enzyme composition and metabolic rates of chemical substances, so direct extrapolation from aquatic to terrestrial toxicity endpoints is hampered. Nonetheless, the literature has stated that aquatic criteria are sufficiently conservative in that terrestrial and soil compartments are sufficiently protected from adverse effects of chemicals (Renaud et al., 2004). However, there are cases reported where substances have been more toxic to terrestrial than aquatic organisms in standardized laboratory studies (Hartmann et al., 2014).

Following the hypotheses and identified uncertainties listed above, the starting point and primary hypothesis of the PROSOIL project were that soil organisms might currently not be sufficiently protected by essential hazard-based criteria. Specific toxicity criteria for CLP & PBT assessments deduced from studies with soil organisms are deemed necessary. A data-driven approach was chosen to derive protective trigger values for soil hazard criteria transparently and, on this basis, derive proposals for soil toxicity criteria for CLP and PBT.

This project aimed to identify substances that were currently not classified as “hazardous for the environment” and those that were not (yet) identified as PBT but possibly should be, based on toxicity criteria derived from studies on soil organisms. It was of significant importance that the proposed toxicity criteria were derived by a data-driven procedure that was transparently documented and reproducible (e.g., if the database changes).

The historical rationales for the definition of the aquatic hazard criteria for classification and labeling of hazardous substances for the environment can be summarized as follows:

- The derivation of aquatic hazard criteria was based on acute toxicity values from algae, fish, daphnids, and other test organisms, if available
- Further criteria concerning biodegradation and bioaccumulation potential were considered to better address the additional hazardous properties.
- The concepts were later extended by chronic toxicity criteria, which defined values based on the calculation of acute-to-chronic ratios.

For the classification and labeling (CLP) of hazardous substances, a similar approach as laid down by Regulation (EC) 1272/2008, more precisely by the provisions under paragraph 4.1.2 for the aquatic environment, could be followed for the derivation of soil hazard criteria.

To derail a toxicity criterion "T" for the PBT assessment, preferably long-term (chronic) indicators of toxicity will be used because recovery cannot be assumed for persistent and bioaccumulative substances.

If a comprehensive analysis of currently available soil data led to a significant number of substances classified as hazardous (CLP assessment) or toxic (PBT assessment) to soil organisms, a comparison between actual and future soil-based classification should be conducted to evaluate the protectiveness of the fundamental approach for the soil compartment.

1.2 Classification of hazardous substances and regulatory data requirements

Classifications of hazardous substances within the CLP regulation and identifying substances as PBT are based on the already available data in the different substance legislations.

Before the European implementation of the classification and labeling system for chemicals according to regulation (EC) No 1272/2008 and the GHS system (Globally Harmonized System for the Classification and Labelling of Chemicals), classifications were regulated by the Dangerous Substance Directive 67/548/EEC of the European Economic Community. This directive regulated the classification and labeling of hazardous substances, e.g., by applying R-phrases (risk) and S-phrases (safety). At that time, the R-phrases held specific aquatic classifications (R50 to R53) and specific risk classifications for terrestrial organisms (e.g., R54 to R57). But, since specific toxicity criteria for terrestrial organisms have never been developed for regulatory purposes, Directive 67/548/EEC relied on the environmental hazard classification from aquatic data only. Since the European CLP Regulation (No. 1272/2008) came into force as the successor of Directive 67/548/EEC, the integration of terrestrial toxicity classifications was entirely excluded and even was not part of the new H-phrases (hazard)- and P-phrases (precautionary) anymore.

While the CLP regulation and its criteria apply to most substances falling under the different substance legislation (except pharmaceuticals), PBT assessments are not harmonized between the various chemical regulations, such as REACH chemicals, plant protection products, biocidal products, and pharmaceuticals. Significant differences in the availability and interpretation of data exist between the regulatory frameworks. For example, the persistence is assessed for REACH chemicals, biocides, and pharmaceuticals by considering temperature-normalized degradation times (to 12°C), while other areas consider the measures under the prevailing laboratory conditions (often 20°C). An attempt to harmonize the PBT assessments has been undertaken and published by the German UBA (Rauert et al., 2014). Also, data requirements on (aquatic) toxicity indicators vary between the pesticides regulation, which has a comprehensive data basis up to (semi-)field studies, and REACH substances that often come with the absolute ecotoxicological “core-data” for low production volume chemicals as data requirements depend on the tonnage manufactured or imported.

Nowadays, REACH, plant protection products, Human and Veterinary Pharmaceuticals, and the Biocidal Products Regulations have data requirements for the toxicity testing of non-aquatic organisms. The REACH regulation only requires information on terrestrial toxicity for chemicals manufactured or imported in quantities > 100 tons/year. However, this requirement may even be omitted if terrestrial exposure is unlikely. The plant protection products regulation (EC) No 1107/2009, which includes the data requirements in Commission Regulation (EU) No 283/2013 for active substances and Commission Regulation (EU) No 284/2013 for formulated plant protection products, requires tests on an additional aquatic invertebrate alongside *Daphnia magna*, honeybees and terrestrial wildlife including mammals, or amphibians for the authorization of new active substances (Hartmann et al., 2014). The Biocidal Products Regulation (EU) No 528/2012 contains information requirements for preparing a dossier for an active substance and subsequent exposure analysis. If, for example, the biocide is applied directly to the soil or released, tests on several non-aquatic species are included.

1.3 Preliminary work on GHS/CLP and PBT soil toxicity criteria

Several authors and working groups expressed concern about the absence of soil toxicity criteria in the past. Concerning CLP/GHS, on the EU level, first discussions between different member states and the work of an OECD expert group lead to a cost-benefit analysis by the EU Commissions Joint Research Center (JRC) and a comprehensive report "cost and benefit analysis of the development and use of environmental hazard-effects classification criteria for terrestrial organisms" (Vega et al., 2003). The report contained several annexes with proposals for terrestrial toxicity criteria developed by different European member states (amongst others: Spain = Carbonell et al., 1997, Nordic countries (Denmark, Finland, Iceland, Norway, Sweden, Greenland, the Faroe Islands, and Åland) = Torstensson et al., 1999, Germany = Feibicke et al. 1999). The review, amongst others, concluded that the development and application of toxicity criteria for terrestrial organisms would beneficially provide more coherence in the European regulatory system. Overall, a terrestrial toxicity criterion would provide complete environmental protection by classifying hazardous substances that escape the (aquatic) classification system.

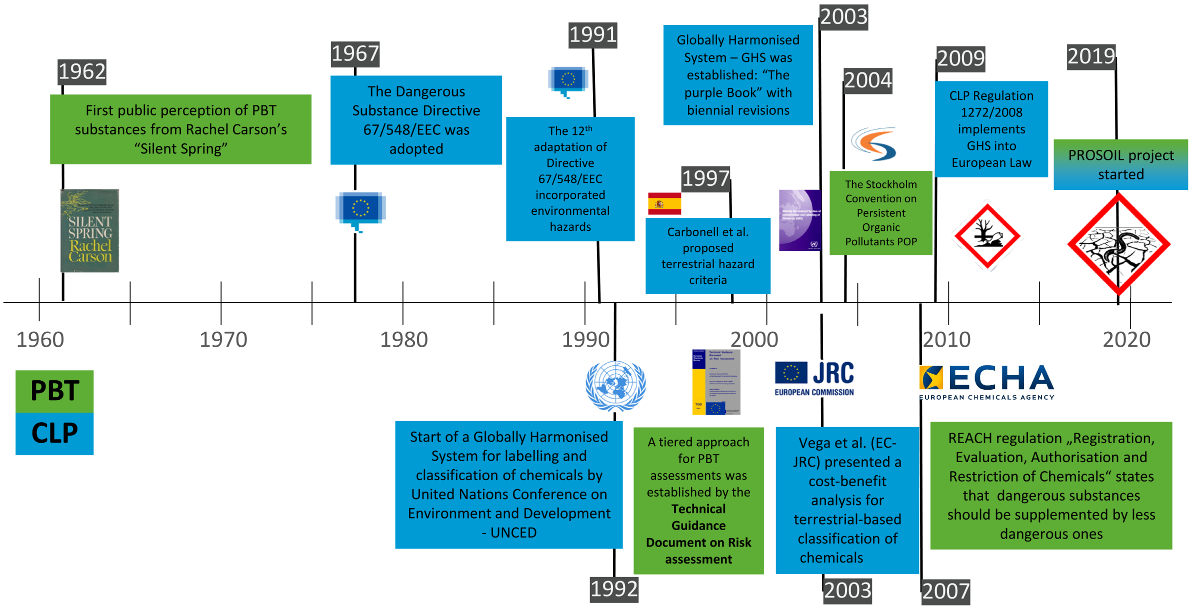
At the same time, in 2002, at the international level, the topic was included in the work plan of the United Nations Sub-committee of the Experts on the Globally Harmonized System of the Classification and Labeling of Chemicals (UN/SCEGHS). During the following years, member states and the OECD expert group brought many proposals and analyses of available approaches and requirements for terrestrial hazard classifications, regularly reported to the UNSCEGHS meetings. This work led to a final and quite comprehensive and concrete proposal at the UN GHS level in 2006 (initiated by Spain) (UN 2006: UN/SCEGHS/12/INF.5). This proposal proposed different hazard categories with criteria for different organism groups. In addition to ecotoxicity values, considerations on biodegradation and bioaccumulation were also included in defining the criteria. Three acute and four chronic classification categories were proposed and defined based on acute (EC_{50}, LD_{50}, or LC_{50}), respectively chronic (NOEC or NOAEL) ecotoxicity data. The categories were established for five different taxonomic groups, considering different exposure routes: microorganisms, in-soil macroorganisms (invertebrates and plants), terrestrial plants considering foliar exposure, and foliar invertebrates/pollinators (focusing on bees), and terrestrial vertebrates (birds and mammals). As a result, it was considered that terrestrial organisms (terrestrial vertebrates, terrestrial plants, soil-ground-foliar dwelling invertebrates, and soil microorganisms) can be exposed to several environmental compartments (soil, air, water, and food). The seven different categories were clearly defined by the respective ecotoxicity criteria for each taxonomic group, with category chronic 4 representing a safety net. In addition, the document included rules for the classification of mixtures as well as proposals for hazard communication which included the symbol hazardous to the environment (categories acute 1 and chronic 1 + 2), a signal word for categories acute and chronic 1 (warning) and hazard statements (e.g., “very toxic to terrestrial life with long-lasting effects” for category chronic 1).

Table 1: Overview of defined hazard criteria in UN 2006.

category	microorganisms	macroorganisms: invertebrates & plants (soil exposure)	terrestrial plants (foliar exposure)	bees	mammals
Acute 1	-	$EC_{50}/LC_{50}/ER_{50}$ ≤ 10 mg ai / kg sdw	EC_{50}/ER_{50} ≤ 10 kg ai / ha	LC_{50} ≤ 1 µg ai / bee	LD_{50} ≤ 5 mg/kg bw
Acute 2	-	$EC_{50}/LC_{50}/ER_{50}$ > 10 but ≤100 mg ai / kg sdw	EC_{50}/ER_{50} > 10 but ≤ 100 kg ai / ha	LC_{50} > 1 but ≤10 µg ai / bee	LD_{50} > 5 but ≤ 50 mg ai / kg bw
Acute 3	-	$EC_{50}/LC_{50}/ER_{50}$ > 100 but ≤1000 mg ai / kg sdw	EC_{50}/ER_{50} > 100 but ≤ 1000 kg ai / ha	LC_{50} > 10 but ≤100 µg ai / bee	LD_{50} > 50 but ≤ 500 mg ai / kg bw
Chronic 1	EC_{50} ≤ 1 mg ai / kg sdw	NOEC ≤ 1 mg ai / kg sdw	-	-	NOAEL ≤ 0.5 mg ai / kg bw
Chronic 2	EC_{50} > 1 but ≤ 10 mg ai / kg sdw	NOEC > 1 but ≤10 mg ai / kg sdw	-	-	NOAEL > 0.5 – 5 mg ai / kg bw
Chronic 3	EC_{50} > 10 but ≤100 mg ai / kg sdw	NOEC > 10 but ≤100 mg ai / kg sdw	-	-	NOAEL > 5 but ≤ 50 mg ai /kg bw
Chronic 4	persistent and/or bioaccumulate (or not rapidly degradable EC_x > 100 mg ai / kg sdw	persistent and/or bioaccumulate (or not rapidly degradable EC_x > 100 mg ai / kg sdw	-	-	persistent and/or bioaccumulative (or not rapidly degradable EC_x > 100 mg ai / kg sdw

In the end, the proposal failed due to political reasons, and the work was not further continued: many member states, as well as the EU Commission, were not entirely convinced of the benefits considering the high efforts that would be needed for implementation – also due to the relatively low data availability at that time.

Figure 1: History of CLP and PBT legislative initiatives.



Source: own illustration, darwin statistics & gaiac

After incorporating the GHS criteria into EU community law, technical assistance was provided to the European Commission. The recommendation was that no hazard classes or toxicity criteria for the terrestrial environment be included in the new CLP system (Ökopol, 2004). However, it was still recommended to continue the examination process.

In their study, Renaud et al. (2004) examined two earlier proposals for terrestrial toxicity criteria in risk assessment (published by Torstensson et al., 1999 and Carbonell et al., 1997) and, because of this, explicitly concentrated on the soil compartment. They reported that the available toxicity data of earthworms (mainly plant protection products) appear comparable to the mentioned proposals for terrestrial toxicity criteria. However, their analysis of the IUCLID database also revealed a lack of appropriate data for classifying terrestrial hazards for most substances, including substances with very high production volumes. And even if the necessary data are available for selected groups of compounds, an extensive experimental test program would be needed to apply a terrestrial classification system to existing substances (Renaud et al. 2004).

Later, discussions on the integration of terrestrial data were initiated in the context of PBT assessment of substances. In 2014 Hartmann et al. published their "Review of available criteria for non-aquatic organisms within PBT/vPvB frameworks" within the JRC Science and Policy reports. The main objective of this review was to gather information and propose approaches for the development of non-aquatic toxicity criteria for the PBT/vPvB assessment. Hartmann et al. stated that criteria for non-aquatic organisms had already been proposed many times and discussed within the United Nations (UN). Unfortunately, they had not been further developed and not implemented in GHS. According to Hartmann et al., hazard data from aquatic systems tended to lead to a more conservative classification, although there were exceptions where higher toxicity to the non-aquatic system can be observed.

In European legislation, chronic aquatic NOEC- or EC_{10}-values of ≤ 0.01 mg/L were applied to each chemical substance to be classified as toxic (T) in the context of PBT assessment. In contrast, for most international legislation and conventions, specific values were $L(E)C_{50}$ ≤ 1 mg/L (acute) and NOEC ≤ 0.1 mg/L (long-term). Therefore, European legislation was less protective from the point of view of PBT classification than other legislations (e.g., US and Canadian legislation).

Hartmann et al. (2014) proposed new toxicity criteria for non-aquatic organisms within the PBT/vPvB assessment framework and how to include them in European legal frameworks. In European legislation, the aquatic toxicity cut-off values for PBT assessment are currently based on the NOEC/EC_{10} values.

Hartmann et al. propose a simplified approach, thus, applying an "L(E)C-to-NOEC extrapolation factor" to establish a relationship between cut-off criteria for hazard classification and corresponding cut-off criteria for toxicity assessment according to PBT/vPvB when establishing terrestrial cut-off values for the PBT assessment of substances. Hartmann et al. also note that the long-term NOEC/EC_{10} value used for PBT assessment under REACH and PPP Regulation would be 100 times lower than the LC_{50} value and ten times lower than the chronic NOEC/EC_x value used for hazard classification in CLP. It is expected that these factors could be used to convert between acute and chronic limits for terrestrial toxicity. However, it is also clearly stated that this simplified approach may need further development.

Finally, in the case of REACH, Hartmann et al. (2014) suggest that data could also be used to investigate the appropriateness of criteria for non-aquatic hazard classification, e.g., under GHS/CLP. In CLP, the aquatic cut-off values for the classification of environmental hazards are based on a combination of chronic NOEC/EC_x values and acute $E(L)C_{50}$ values. Depending on the degradability of the substance, different chronic NOEC or EC_x values are used as cut-off values within CLP (category 1 to 3). The lowest value (Aquatic Chronic Category 1: NOEC/$EC_x \leq 0.01$ mg/L) corresponds to the toxicity criteria for identifying PBT substances under the REACH regulation. If sufficient chronic toxicity data are not available, a substance may also be classified as chronic category 1 based on $L(E)C_{50} < 1$ mg/L in combination with non-rapid degradability and BCF ≥ 500 (Hartmann et al. 2014).

Regarding further developing terrestrial toxicity criteria in PBT and CLP assessments, Hartmann et al. remark that there are often disagreements about appropriate values in scientific literature, international legislation, and official reports. The differences within specific organism groups and exposure routes are up to a factor of 60, generally up to a factor of 10.

Table 2: Suggested Terrestrial toxicity criteria as they can be found in previous work and literature. Most of the literature values are divided into three hazard categories in descending order: Very toxic/Acute Category 1/Chronic Category 1 (Cat 1), Toxic/Acute Category 2/Chronic Category 2 (Cat 2), Harmful/Acute Category 3/Chronic Category 3 (Cat 3). Data from US EPA, 2012 have been partially extended by the additional category: slightly toxic (Cat 4). (dw = dry weight, bw = body weight).

Reference	Organism	Cat 1	Cat 2	Cat 3	Cat 4
Carbonell et al., 1997 (Spanish proposal in Vega et al., 2003)	Earthworms	$EC_{50} < 1$ mg/kg soil (dw)	$EC_{50} = 1$-10 mg/kg soil (dw)	$EC_{50} = 10$-100 mg/kg soil (dw)	
	Bees	$LD_{50} \leq 1$ µg/bee	$LD_{50} = 1$-10 µg/bee	$LD_{50} = 10$-100 µg/bee	
	Microorganisms	$EC_{50} \leq 1$ mg/kg soil (dw)	$EC_{30\text{-}50} \leq 1$ mg/kg soil (dw)		
	Terrestrial plants (soil exposure)	$EC_{50} \leq 1$ mg/kg soil (dw)	$EC_{50} = 1$-10 mg/kg soil (dw)	$EC_{50} = 10$-100 mg/kg soil (dw)	

Reference	Organism	Cat 1	Cat 2	Cat 3	Cat 4
	Terrestrial plants (foliar exposure) (biomass or seed germination)	EC_{50} ≤ 1 kg/ha	EC_{50} = 1-10 kg/ha	EC_{50} = 10-100 kg/ha	
	Birds and mammals (oral exposure)	LD_{50} ≤ 25 mg/kg (bw)	LD_{50} = 25-200 mg/kg (bw)	LD_{50} = 200-2000 mg/kg (bw)	
	Terrestrial vertebrates (air exposure)	LC_{50},4h ≤ 0.5 mg/L (air)	LC_{50},4h = 0.5-2 mg/L (air)		
Feibicke et al., 1999 (German proposal in Vega et al., 2003)	Soil dwelling invertebrates (springtails, earthworms; soil exposure)	EC_{50} ≤ 10 mg/kg soil (dw)	EC_{50} = 10-100 mg/kg soil (dw)		
	Bees	EC_{50} ≤ 10 µg/bee	EC_{50} = 10-100 µg/bee		
	Terrestrial plants (soil exposure)	EC50 ≤ 10 mg/kg soil (dw)	EC_{50} = 10-100 mg/kg soil (dw)		
	Birds	EC_{50} ≤ 25 mg/kg (bw)	EC_{50} = 25-200 mg/kg (bw)		
Torstensson et al., 1999 (Nordic proposal in Vega et al., 2003)	Soil dwelling invertebrates including earthworms	EC_{50} ≤ 10 mg/kg soil (dw)	EC_{50} = 10-100 mg/kg soil (dw)		
Renaud et al., 2004	Earthworms (classification method C3)	EC_{50} < 4 mg/kg (dw)			
	Earthworms (classification method C4)	EC_{50} < 60 mg/kg (dw)			
ECB, 2000	Acute toxicity for soil-dwelling organisms (including earthworms/ plants)	$L(E)C_{50}$ ≤ 10 mg/kg (dw)	$L(E)C_{50}$ = 10-100 mg/kg (dw)	$L(E)C_{50}$ = 100-1000 mg/kg (dw)	
Hartmann et al., 2014	Soil dwelling invertebrates incl. earthworms (soil exposure)	≤ 0.01 to ≤ 0.6 mg/Kg (dw)	≤ 0.1 to ≤ 6 mg/Kg (dw)	≤ 0.1 mg/Kg	
	Foliar invertebrates and pollinators (incl. bees)	≤ 0.01 to 0.02 µg/bee	≤ 0.1 to 0.2 µg/bee		
	Microorganisms	≤ 0.01 mg/Kg (dw)	≤ 0.1 mg/Kg (dw)		
	Terrestrial plants (soil exposure)	≤ 0.01 to ≤ 0.1 mg/kg dw	≤ 0.1 to ≤ 1 mg/kg dw		
	Terrestrial plants (foliar exposure)	≤ 0.01 to ≤ 0.1 kg/ha	≤ 0.1 to ≤ 1 kg/ha		

Reference	Organism	Cat 1	Cat 2	Cat 3	Cat 4
	Terrestrial vertebrates (birds and mammals) (oral exposure)	≤ 0.05 to ≤ 0.5 mg/kg bw (≤ 0.4 to ≤ 4 ppm in food for chicken; ≤ 1 to ≤ 10 ppm in food for rat)	≤ 0.5 to ≤ 5 mg/kg bw (≤ 4 to ≤ 40 ppm in food for chicken; ≤ 10 to ≤ 100 ppm in food for rat)	≤ 0.05 mg/Kg bw (≤ 0.4 ppm in food for chicken; ≤ 1 ppm in food for rat)	
	Terrestrial vertebrates (birds and mammals) (diet exposure)	≤ 5 ppm in food	≤ 50 ppm in food		
	Terrestrial vertebrates (air exposure)	≤ 0.05 mg/L air	≤ 0.5 mg/L air		
NZ EPA, 2012	Earthworms	EC_{50} < 1 mg/kg soil (dw)	EC_{50} = 1-10 mg/kg soil (dw)	EC_{50} = 10-100 mg/kg soil (dw)	
	Terrestrial invertebrates in general	LD_{50} ≤ 2 µg/animal	LD_{50} = 2-11 µg/animal	LD_{50} = 11-25 µg/animal	
	Microorganisms	EC_{50} ≤ 1 mg/kg soil (dw)	EC_{50} = 1-10 mg/kg soil (dw)	EC_{50} = 10-100 mg/kg soil (dw)	
	Terrestrial vertebrates (birds and mammals; oral exposure)	LD_{50} ≤ 50 mg/kg (bw)	LD_{50} = 50-500 mg/kg (bw)	LD_{50} = 500-2000 mg/kg (bw)	
	Terrestrial vertebrates (birds and mammals; diet exposure)	LC_{50} ≤ 500 ppm in the diet	LC_{50} = 500-1000 ppm in the diet	LC_{50} = 1000-5000 ppm in the diet	
US EPA, 2012	Bees	LD_{50} ≤ 2 µg/bee	LD_{50} = 2-11 µg/bee	LD_{50} > 25 µg/bee	
	Terrestrial vertebrates (birds and mammals; oral exposure)	LD_{50} ≤ 10 mg/kg (very highly toxic)	LD_{50} = 10-50 mg/kg (highly toxic)	LD_{50} = 50-500 mg/kg (moderately toxic)	LD_{50} = 500-5000 mg/kg (slightly toxic)
	Terrestrial vertebrates (birds and mammals; diet exposure)	LD_{50} ≤ 50 ppm (very highly toxic)	LD_{50} = 50-500 ppm (highly toxic)	LD_{50} = 500-1000 ppm (moderately toxic)	LD_{50} = 1000-5000 ppm (slightly toxic)

Overall, all preliminary work showed that the development of terrestrial toxicity criteria would significantly strengthen protection levels from environmental hazards. It was often stated that it appeared incoherently in the regulatory practice to pretend the protection of all environmental compartments, the classification, and often also the authorization of chemicals, exclusively on aquatic toxicity data.

1.4 Definition of in-soil organisms

The risk assessment procedures and concepts that are implemented nowadays were developed in the history of about 50 years of the development of ecotoxicological experimental methods. There is still a tradition to draw a line between aquatic and terrestrial organisms and thus also between aquatic and terrestrial study types. Aquatic taxa comprise crustaceans, higher aquatic plants, algae, vertebrate fish, and insects. Terrestrial organism groups comprise birds, mammals, invertebrates, higher plants, and arthropods, mainly insects. In a broader sense, the ECHA Technical Guidance Document states in its in-depth guidance for hazard assessments (ECHA, 2017, chapter R7c) that terrestrial organisms may be exposed to substances released into the environment. Thus, ecosystem functions and processes may be affected negatively.

It recognizes the outstanding complexity of the terrestrial compartment (compared to, e.g., aquatic ecosystems): *"...a comprehensive effect assessment for the whole compartment can only be achieved by a set of assessment endpoints covering (i) the different routes by which terrestrial organisms may be exposed to substances (i.e., air, food, pore water, bulk-soil) and (ii) the most relevant taxonomic and functional groups of terrestrial organisms (microorganism, plants, invertebrates, vertebrates)...restricted to soil organisms in a narrow sense, i.e., on non-vertebrate organisms living the majority of their lifetime within the soil and being exposed to substances via the soil pathway…".*

The assessment should be done in line with common practices in ERA for various substance legislation in the EU. Even though terrestrial ecosystems comprise above-ground and groundwater communities, here only effects on soil organisms are addressed that are exposed via soil particles or soil pore water (according to the European Communities 2003), i.e., terrestrial invertebrates living above-ground (e.g., ground-dwelling beetles), terrestrial vertebrates living a part of their lifetime in soils (e.g., mice), groundwater organisms (invertebrates and microorganism), and indirect adverse effects on soil functions are not included in the group of "in-soil" organisms.

For the PROSOIL project, it was decided that in-soil organisms should comprise the following groups – following the above-outlined definitions:

- Plants exposed via soil
- Microorganisms directly affected in their functional roles in soil ecosystems
- Macroorganisms from seven taxonomic groups (mites, enchytraeids, nematodes, lumbricids, isopods, myriapods, and collembolans) and the unique group of dung-insects

1.5 Regulatory framework – relevant directives and guidelines

Terrestrial toxicity data are required in the current chemicals' regulation, depending on their specific applications. Plant protection products are intentionally applied to soils and designed to harm specific target organisms, which leads to more comprehensive data requirements for aquatic and terrestrial data than for unintentionally released chemicals. The available data is the basis for classifications under CLP. Therefore, the contents and terrestrial references of the most critical European chemical regulations will be discussed in more detail in the following section. The principal regulations and directives to be considered in this project are the following

Chemicals

- Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorization, and Restriction of Chemicals (REACH)

Plant Protection Products (divided into active substances and their metabolites, formulations)

- Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market
- Commission Regulation (EU) No 283/2013 of 1 March 2013 setting out the data requirements for active substances, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and the Council concerning the placing of plant protection products on the market
- Commission Regulation (EU) No 284/2013 of 1 March 2013 setting out the data requirements for plant protection products, in accordance with Regulation (EC) No 1107/2009 of the European Parliament and the Council concerning the placing of plant protection products on the market

Biocidal products

- Regulation (EU) No 528/2012 of the European Parliament and of the Council of 22 May 2012 concerning the making available on the market and use of biocidal products

Pharmaceuticals (human use and veterinary use)

- Regulation (EC) No 726/2004 of the European Parliament and of the Council of 31 March 2004 laying down Community procedures for the authorization and supervision of medicinal products for human and veterinary use
- Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use
- Guideline on the environmental risk assessment of medicinal products for human use, 01 June 2006, Doc. Ref. EMEA/CHMP/SWP/4447/00 corr 2
- Directive 2004/28/EC of the European Parliament and of the Council of 31 March 2004 amending Directive 2001/82/EC on the Community code relating to veterinary medicinal products
- Guideline on the environmental impact assessment for veterinary medicinal products phase II, CVMP/VICH/790/03-FINAL

Of these regulations and directives, the Commission Regulations (EU) No 283/2013 for active substances of plant protection products and No 284/2013 for setting out the data requirements for plant protection products, both within Regulation (EC) No 1107/2009, require mandatory terrestrial toxicity data for the approval of PPP. Most other directives require these data for substances to surpass certain trigger values or specific exposure situations. Under the REACH regulation, only substances with a production volume of >100 t/a, or in case of adsorption to soil, persistency, or specific hazardous properties, or if exposure to soil cannot be excluded require terrestrial data for registration. To determine the specific data requirements for a substance, so-called soil hazard categories 1 to 4 are applied, referring to the physicochemical properties and using the “equilibrium partitioning method” (EPM). The situation for soil toxicity studies is similar to Regulation (EU) No 528/2012 regarding biocidal products. Here, tests are required if the risk assessment for the terrestrial compartment (based on the equilibrium partitioning method) indicates a concern or if there is direct or long-term exposure. However, tests on terrestrial organisms will be mandatory in the core data set for some biocidal product types. In the case of Directive 2001/83/EC on medicinal products for human use within the Regulation (EC) No 726/2004, the K_{OC} (> 10000 L/kg) is decisive for the submission of terrestrial data unless the substance of concern is readily biodegradable. In Directive 2004/28/EC of the European Parliament and of the Council of 31 March 2004 amending Directive 2001/82/EC on veterinary medicinal products, the submission of terrestrial toxicity data is only applicable if the chemical is used with direct soil exposure or by spreading of manure on pasture.

Overall, European chemical regulation is currently much more based on aquatic data than terrestrial data. This focus, of course, impacts the data availability and significantly restricts it compared to the availability of aquatic toxicity data. Table 3 shows the toxicity data and guidelines required by the various regulations and directives for a better overview.

Table 3: European regulatory data requirements for the terrestrial compartment focus on the relevant taxa considered in this project and the most important soil-related test guidelines used. Data requirements are shown for Regulation (EC) No 1907/2006 (REACH) according to Annex IX (production volumes 100-1000 t/a) and (Annex X) (production volumes ≥ 1000 t/a)), Regulation (EC) No 1107/2009 for active substances (PPP active), formulations (PPP form), and metabolites (PPP meta) in plant protection products, Regulation (EU) No 528/2012 for biocidal products (Biocides), and Regulation (EC) No 726/2004 for medicinal products for human (Med human) and veterinary products (Med vet). For REACH, tests may be waived when exposure to soil can be excluded. For PPP, "individual" means that additional data can be waived if available data for a.i. is relevant. For biocides, "individual" means that data requirement depends on product type, direct exposure to soil, or identified concern for the terrestrial compartment.

Organism group	Type of test	Guideline	REACH 100 – 1000 t/a (A. IX)	REACH ≥ 1000 t/a (A. X)	PPP active	PPP form	PPP meta	Biocides	Med human	Med vet
Soil-dwelling invertebrates	Earthworm acute	OECD 207; EC method C.8; ISO 11268-1; ISO 11268-2	Short-term (long-term in case of persistency or adsorption potential)	-	-	-	-	individual	OECD 207 mandatory (Phase II Tier B) or EC C.8	-
Soil-dwelling invertebrates	Earthworm Subacute/ Reproduction	OECD 222; ISO 11268-1		Long-term	mandatory	individual	individual	individual	-	mandatory (Phase II Tier B)
Soil-dwelling invertebrates	Nematode Caenorhabditis elegans	ASTM E2172	-	-	-	-	-	individual	-	-
Soil-dwelling invertebrates	Enchytraeid Reproduction Test	OECD 220; ISO 16387	-	-	-	-	-	individual	-	OECD 220 mandatory (Phase II Tier B)
Soil-dwelling invertebrates	Predatory mite Hypoaspis	OECD 226	-	-	mandatory	individual	individual	individual	-	-
Soil-dwelling invertebrates	Collembolan reproduction	OECD 232; ISO 11267	OECD 232 or ISO 11267 (see requirements)	OECD 232 or ISO 11267 (see requirements)	OECD 232 mandatory	OECD 232 individual	OECD 232 individual	individual	ISO 11267 mandato	-

Organism group	Type of test	Guideline	REACH 100 – 1000 t/a (A. IX)	REACH ≥ 1000 t/a (A. X)	PPP active	PPP form	PPP meta	Biocides	Med human	Med vet
									ry (Phase II Tier B)	
Arthropods	Side-effects of PPP on non-target arthropods	IOBC, 2000	-	-	mandatory	individual	-	individual	-	-
Dung organisms	Dung fly larvae, Aphodius constans/ Onthophagus taurus	OECD 228	-	-	-	-	-	-	-	mandatory for parasiticide s
Dung organisms	Dung beetle larvae, Scathophaga stercoraria, Musca autumnalis	OECD GD 122	-	-	-	-	-	-	-	mandatory for parasiticide s
Plants	Plants seedling emergence	OECD 208	Short-term (long-term in case of persistency or adsorption potential)	Long-term	individual	individual	-	individual	-	mandatory (Phase II Tier A)
Plants	Plants seedling growth	OECD 208; ISO 22030	Short-term (long-term in case of persistency or adsorption potential)	Long-term	-	-	-	individual	OECD 208 mandato ry (Phase II Tier B)	OECD 208 mandatory (Phase II Tier A)
Plants	Vegetative vigor	OECD 227	Short-term (long-term in case of persistency or adsorption potential)	Long-term	individual	individual	-	individual	-	-

Organism group	Type of test	Guideline	REACH 100 – 1000 t/a (A. IX)	REACH ≥ 1000 t/a (A. X)	PPP active	PPP form	PPP meta	Biocides	Med human	Med vet
Microorganisms	Soil-microorganisms N-transformation	OECD 216; EC method C.21	mandatory	mandatory	OECD 216 mandatory	individual	individual	individual	OECD 216 mandatory (Phase II Tier B)	OECD 216 mandatory (Phase II Tier A)
Microorganisms	Soil-microorganisms C-transformation	OECD 217; EC method C.22	-	-	-	-	-	individual	individual	individual
Microorganisms	N-mineralization and nitrification	ISO 14238:2012	-	-	-	-	-	individual	individual	individual
Microorganisms	Activity of the soil microflora	BBA Part VI, 1.1	-	-	-	-	-	individual	individual	individual
Microorganisms	Dehydrogenase activity in soils	DIN EN ISO 23753-2	-	-	-	-	-	individual	-	-

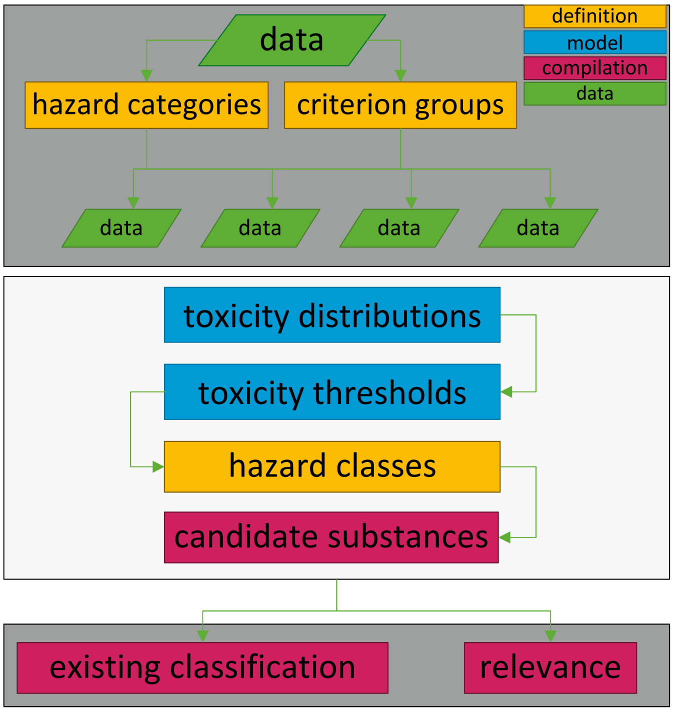
1.6 The PROSOIL conceptual approach

The project aimed to establish a methodology suited to demonstrate whether **soil organisms are sufficiently protected by the present CLP & PBT assessments.** A sound and transparent scientific basis was considered indispensable to ensure that the newly involved toxicity criteria for soil organisms would be trusted by the responsible regulatory bodies and could reliably be communicated to the broader public and other stakeholders.

The main aims were

- the **derivation of toxicity thresholds** as a basis to assign toxicity criteria (i.e., harmful, toxic, very toxic) for specific hazard categories (e.g., acute, or chronic toxicity to soil organisms),
- the contribution to **proposals of toxicity criteria** for classification & labeling and
- **contributions to defining a soil-based T-criterion** for the PBT assessment (using the identified toxicity thresholds).

Figure 2: Representation of the conceptual approach and workflow of the PROSOIL project.



Source: own illustration, darwin statistics & gaiac

The PROSOIL project contributes to protecting soil organisms in chemical regulatory frameworks by compiling a comprehensive, harmonized database of ecotoxicity data for soil organisms from several international regulatory data repositories (for the first time in such a comprehensive manner). Moreover, candidate lists of chemicals from data-based threshold ecotoxicological values were derived. This exercise identified potential hazards from (registered) chemicals that were not sufficiently covered by the existing aquatic toxicity criteria (a schematic overview is given in Figure 2).

Urgent needs and consequences were formulated to address future adaptations of the existing regulations.

Definition of hazard classification under CLP

The **HAZARD CLASS** reflects physical, health, environmental and additional hazards, e.g., “hazardous to the aquatic environment”.

Different **HAZARD CATEGORIES** exist for the different hazard classes, e.g., long-term or short-term toxicity for aquatic organisms, differentiated for different categories and toxicity ranges, e.g., chronic 1-3 and acute 1-3.

HAZARD CRITERIA are defined for each category to have a relative measure of the toxic hazard of a substance, e.g., if it is deemed “very toxic“, “toxic“, “harmful“, or “non-toxic“ based on defined threshold values for ecotoxicity data.

Hazard categories and criteria are established for different taxonomic groups (e.g., invertebrates, algae, fish for the aquatic environment).

HAZARD STATEMENTS, e.g., *“HS412 - Very toxic to life in soils with long-lasting effects,”* and **LABELS** (pictograms) are used to communicate the hazards in a defined way.

This report summarizes the main rationales, approaches, and results of the PROSOIL-project. However, the work in the background on the interoperability, on the common ecotoxicological ontologies of very diverse data repositories from scratch by profound and expert elicitation, was much more excellent than the merely reported text could ever express the efforts accomplished. More than 15.000 single lines of R-code (admittedly including functions equivalent to dynamic link libraries) were distributed over 65 single R-scripts and imported, processed, and produced 270 data tables. It filled 436 pages of (executable) text if the code was printed out.

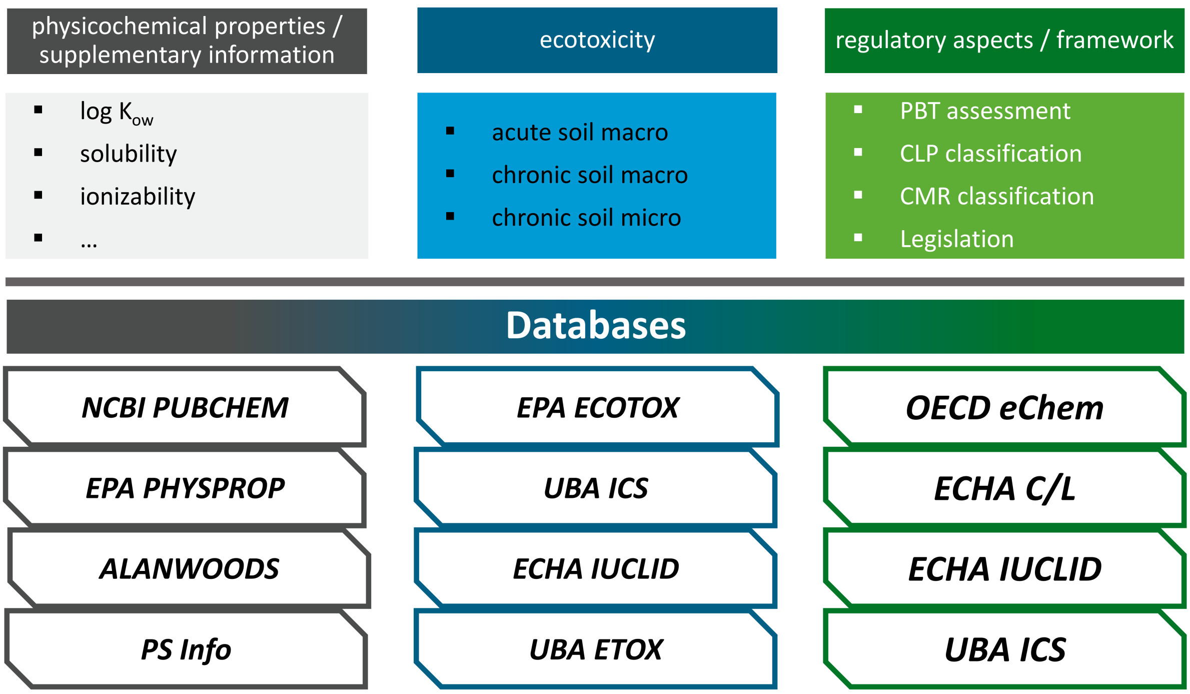
2 Structure of the ecotoxicological effect database

First, substance identifiers and ecotoxicological data for soil organisms (including endpoints, test conditions, exposure pattern, duration, and others) were needed for developing ecotoxicological hazard criteria for chemicals. Furthermore, additional data were needed for a later comparison of the developed hazard criteria with existing classifications, such as information on the actual regulatory status and existing provisions of a substance (CLP, CMR, SVHC, PBT status, Industrial Emissions Directive 2001/75/EU, Water Framework Directive 2000/60/EC). Additionally, physicochemical properties of the chemicals were collected for potential further analyses.

This information can be gathered through different governmental, data-protected repositories, and publicly available databases. Figure 3 describes the linkage between the necessary information and the respective databases. The essential data are shown in the upper part, arranged by the information types. The lower part shows the currently used databases for data extraction for this project.

The data compilation performed in this project mainly focused on study data from soil organisms, i.e., earthworms, collembolans, mites, fungi, bacteria (often potentially summarized as "microorganisms"), as well as plants and non-target arthropods if the test guideline directly intends exposure via soil and reliably represented by the test method. Acute and chronic data were considered equivalently valuable but were separated in a later step. In general, it was considered essential to first look at the data availability before decisions on the focused organism groups would be made. According to known risk assessment schemes, and depending on data availability, the inclusion of other organism groups was considered and intensively discussed. For instance, it was decided to include (wild)-bees and litter-dwelling organisms in the overall database because of their high relevance for the discussions on massive insect declines and indirect effects.

Figure 3: Information on effects, properties, and regulatory classification data of active substances.



Source: own illustration, darwin statistics & gaiac

Generally, not included were toxicity data on birds and mammals since there was no direct exposure via soil to be expected (indirect effects via secondary poisoning, an approach that involves many assumptions and thus uncertainties) and the simulated exposure routes in the respective (clinical) studies are not adequately transferable to soil concentrations.

Four comprehensive ecotoxicological datasets were queried (between 2019 and 2020) and processed to achieve a self-contained, harmonized, and consistent database across chemical regulations and regulatory assessment areas. The intention was that the database should contain the most significant possible part of all the ecotoxicological information with soil organisms available. Many data repositories were open for the public and made available via web portals, which was considered an essential step towards increased transparency and was also the favored approach for this project. However, since regulatory authorities and companies (depending on the regulatory area) often use non-publicly available data, those confidential, data-protected standard data were also included in the database. These datasets represented the basis on which decisions for CLP and PBT assessments were taken.

The four databases utilized in this project were the following: the ICS- and ETOX-system, both hosted by the German Environment Agency, the IUCLID database hosted by ECHA and OECD, and the ECOTOX data repository of the US EPA. These repositories have been considered to compile data on the environmental toxicity of chemicals. Additional data on physicochemical properties and the regulatory status have been added if the substance has been found in the databases mentioned above, pre-processed, and added as a data entry. A variety of databases offer open access to this complementary information, such as the NIH PubChem-, the RSC Chemspider-, and the SRC-PHYSPROP-databases. Furthermore, regulatory relevant physicochemical and regulatory data on notification status and classification and labeling were also available from ICS and IUCLID.

2.1 Ecotoxicological databases

The query and transfer of large databases mean substantial efforts that cannot be easily repeated regularly. However, the process of data compiling is organized along a workflow of automated scripts that provide the opportunity to repeat the query for updated data at a later stage.

2.1.1 UBA ICS

The German Environment Agency – UBA maintained a data repository "Informationssystem Chemikaliensicherheit – Information Systems on Chemical Safety - ICS" for relevant information submitted by registrants of chemicals regulated under German and EC laws. Two regulatory frameworks fed the database with validated ecotoxicological study information: Data submitted under the German Medicines Act (Arzneimittelgesetz – AMG) and the Plant Protection Act (Pflanzenschutzgesetz – PflSchG). UBA section "IV 2.1 – Information Systems on Chemical Safety" was responsible for maintenance and further development. Information on REACH chemicals and biocidal substances (besides identity information) and data on the fate and behavior of chemicals were not included. Data on REACH chemicals and biocidal products were collected and maintained in the IUCLID system but not validated by authorities (refer to section 2.1.4). The database held data on the identity and ecotoxicological effects of products, active substances, and metabolites. Authorized governmental officers fed data into the database via an "update module", containing picklists, free text fields, and with the help of standard operating procedures (SOP) that set the benchmarks of validity and completeness of the data. SOPs were constantly subjected to revision and exhibited differences between regulatory areas. The database system looked back on more than 30 years of usage, maintenance, extension, and development, which brought a considerable ambiguity in terminologies that needed to be handled and aligned within and between the different databases.

Data were queried with the help of the specialists of the chemicals identity management unit that were in charge of the ICS system at UBA. It was made available in a single long-version "cross-table" contained in a Microsoft Access database and held all information available. The data was then placed in a data-processing pipeline described in chapter 3 in detail.

2.1.2 UBA ETOX

The „Informationssystem Ökotoxikologie und Umweltqualitätsziele" ("Information System Ecotoxicology and Environmental Quality Targets - ETOX") of the German Environment Agency contained data on validated ecotoxicological test results and environmental quality standards that were publicly accessible via the internet (Schudoma et al., 2008). It held 55.000 datasets on the effects of chemicals on aquatic organisms, 5.000 datasets for terrestrial organisms, and 4.500 datasets on environmental quality standards for water and soil (https://webetox.uba.de/webETOX/index.do, accessed 2019-10-04).

2.1.3 US EPA ECOTOX

The Environmental Protection Agency of the United States of America – US EPA maintained the "ECOTOXicology knowledgebase" (US EPA - United States Environmental Protection Agency 2018), a compilation of toxicity data from the areas of aquatic organisms, terrestrial plants, and other terrestrial species, originally distributed over three independent repositories (AQUIRE, PHYTOTOX, TERRETOX). From the year 1981 onwards, the Agency compiled mainly data from the publicly available, peer-reviewed literature, which led to more heterogeneous input data and formats, but also to a greater variety of tested species and experimental designs compared to the databases described above that were merely a set-up for regulatory purposes (with a limited set of internationally well-standardized studies). The ECOTOX-database was particularly interesting in obtaining supplementary data on rarely tested chemicals and species combinations. It could be queried via the web interface at https://cfpub.epa.gov/ecotox/ which offered numerous filter functions and export options. Complete dumps of the database tables were available as pipe (|)-separated files and could be reassembled by lookup tables with primary and foreign keys.

2.1.4 ECHA IUCLID

The International Uniform Chemical Information Database - IUCLID was available in version 6. It provided information on the identity and intrinsic properties, the classification of environmental hazards and labeling, and the ecotoxicological effects of chemical substances within a software environment (European Chemicals Agency, 2018). It aimed at facilitating the storage and exchange of data for the chemical industry and the responsible legal authorities to comply with the respective legislation of placing chemicals on the market. Therefore, mainly data required under the REACH regulation were included. The system was developed and maintained by the European Chemicals Agency (ECHA) in Cooperation with the Organisation for Economic Co-operation and Development (OECD) and the European Commission. The major advance and benefit lay in using standardized "OECD Harmonized Templates" provided in an XML format (eXtensible Markup Language) that facilitated the industry to fulfill their obligations. Standardization also facilitated the information extended to biocidal or plant protection product information requirements. Data often represented a single dossier from one registrant for a single chemical substance or grouped information from dossiers of multiple registrants, serving as the basis for the output of risk assessments and reports, i.e., the "Chemicals Safety Assessment" or the "Chemicals Safety Report". Data could be imported as i6z-files to desktop or server versions of the IUCLID software. REACH data were not subject to regular quality checks (e.g., for validity, reliability, plausibility), but in a random and concern-based targeted process, compliance checks ensured the completeness and appropriateness of data or proposed testing procedures. Regulators work downstream of IUCLID on the hazard classifications and restrictions - information not included in IUCLID.

The approach which offered the highest degree of self-sufficiency was the download of all REACH study results in a single dump of the Oracle database, which held the complete structure of the IUCLID system. As the support of the open-source database “PostgreSQL” was deprecated with IUCLID5, a local installation of the commercial software Oracle Database 18c Express was executed and fully integrated into the computer's operating system. The R-package “RODBC” offered connectivity to ODBC databases and enabled the data transfer to the unified, project-related PostgreSQL-database system (Ripley & Lapsley, 2017).

Data on REACH study results with the official IUCLID6-application were downloaded from the IUCLID-website (https://iuclid6.echa.europa.EU/REACH-study-results). The data package contained information on 15.000 substances in a single IUCLID database and could then be accessed locally after import into the desktop version of IUCLID6. Data could be somehow “simplified, " i.e., deleted confidential information on exposure or intended uses. REACH study results seemingly date to 2017-03-03. The desktop software was used to ensure the correctness and plausibility of the data compilation from the Oracle database dump.

OECD offered several documents that described the general structure of IUCLID dossiers submitted, e.g., biocides, plant protection products, and REACH substances. However, the compilation and reconstruction from the “raw” database dump required substantial efforts and some support from the ECHA helpdesk.

2.2 Physicochemical properties databases

2.2.1 UBA ICS

The UBA ICS database was described in detail in chapter 2.1.1. Besides data on the identity and ecotoxicological effects of products, active substances, and metabolites, the UBA ICS database also held information on the physicochemical properties of the respective substance. This information was extracted separately and added to the data compilation during the physicochemical properties import.

2.2.2 NIH PUBCHEM

From the web address https://pubchem.ncbi.nlm.nih.gov/, the interface of the PubChem database was available. With approximately 103 million listed compounds, PubChem was one of the largest databases for physicochemical information currently publicly available. It belonged to the US National Institute of Health (NIH). The database held information relevant to chemical regulations, such as identities, physicochemical properties, chemical safety, and toxicity. Systematic Web Scraping was not foreseen; however, a list of chemical identifiers could be fed into the data portal, which enabled bulk retrieval of information. Lists could be generated using the R-package webchem by the function get_cid. The list could then be passed via the PUG-REST API (Power User Gateway to Representation State Transfer via an Application Programming Interface) to the functions “pc_prop” of the R-package “webchem” (Szöcs et al., 2015) to retrieve chemical properties or “pc_synonyms” to find synonyms of PubChem entries. The data source will be primarily used to get information on the chemical properties of active substances.

2.3 Regulatory status databases

ECHA provided and published lists and further documents with the results of the applicants' assessments (CLP, PBT) and for hazardous substances of particular concern, regulated under REACH or the Biocidal Products Regulation, of the formalized Harmonized Classification and Labelling procedure - CLH done by the MS and ECHA (represented by the Committee for Risk Assessment – RAC) on their websites. Ongoing assessments for PBT substances were listed at https://echa.europa.EU/de/PBT. Information on classification and labeling for notified substances from manufacturers and importers, e.g., H-phrases, is provided by the CL inventory of ECHA (https://echa.europa.EU/de/information-on-chemicals/cl-inventory-database). The first classification results under CLP of REACH substances that were carcinogenic, mutagenic, or toxic for reproduction could be obtained from http://echa.europa.EU/documents/10162/13562/cmr_report_en.pdf. REACH substances identified as PBT according to Article 57f were listed on the candidate list for SVHC substances. Substances that had undergone a harmonized classification and labeling procedure and were classified as hazardous (after the decision of ECHAs risk assessment committee (RAC) and the EU Commission) were included in Table 3 of Annex VI to the CLP Regulation and were available in an updated list via ECHA website (https://echa.europa.eu/de/information-on-chemicals/annex-vi-to-clp) and were contained as well in the C&L inventory of ECHA. Relevant information and assessment reports for biocidal active substances were provided on the website: https://echa.europa.EU/information-on-chemicals/biocidal-active-substances.

The EU chemicals legislation finder (EUCLEF) provided an overview of all relevant regulations for a single chemical substance. (https://echa.europa.eu/de/information-on-chemicals/EUCLEF): "EUCLEF is an easy-to-use online service that gives you free access to a comprehensive overview of EU chemicals legislation in one place. It is integrated with our chemicals database and makes it possible for you to search for a substance, find out how it is regulated under different pieces of legislation and what obligations you have."

The European Commission provided comprehensive information on the regulatory status of plant protection products under https://ec.europa.eu/food/plant/pesticides/eu-pesticides-database/.

Tables of the regulatory status of a substance that was notified for use as a human or veterinary medicine could be downloaded from the website of the European Medicines Agency – EMA (https://www.ema.europa.eu/en/medicines/download-medicine-data).

3 Data process structure

As described above, the data used in the present study originated from various sources and depicted various levels of data quality. Data was not necessarily validated by public authorities, particularly for REACH chemicals, where companies submitted data without official approval. Hence data was of varying quality. Also, important to consider, was that most ecotoxicological, physicochemical properties, and even regulatory databases used the active substance as the unique atomic unit. Product names were not provided except by the ICS database. To match entries from different organism groups and exposure scenarios and avoid the trap of mixture toxicities from combination products, the active substance was addressed by the internationally agreed unique identifiers CAS numbers or by IUPAC names. It was aimed at a comprehensive and unified data analysis that led to a generic derivation of toxicity hazard criteria across regulatory frameworks. This aim required a fully comprehensible and reproducible data processing pipeline, which was outlined as a workflow of data compilation, quality assurance, and explorative data analysis. A workflow for the compilation and processing of data was documented in the following. It described how data was imported into the project database and joined for further use. The workflow comprised steps of data retrieval from the identified data repositories, data pre-processing to harmonize attributes (e.g., variable/column names), units (e.g., soil chemical concentrations, test durations), and nomenclatures (e.g., test species, measurement endpoints). An exploratory data analysis was the subject of the following chapters and was not reported here. The latter should help define filters (data quality) and groups (comparable subsets as “criterion groups”) to create a solid basis for downstream work packages that aim to deduce soil toxicity criteria.

Three variables served as unique identifiers of the chemical substances: the common name “*aicommon*”, the CAS number “*aicas*”, and the IUPAC name “*aiiupac*” (systematic nomenclature entry of the International Union of Pure and Applied Chemistry). Furthermore, the datasets had to contain sixteen ecotoxicological core variables. These core variables were as follows:

testid	Identification number within the source database
aicommon	Common name
aicas	CAS-number
aiiupac	IUPAC name
testspecies	Test species
origin	Name of the source database
testtype	Type of the test (e.g., laboratory test, field test
measendpoint	Measurement endpoint (e.g., mortality, emergence)
statendpoint	Statistical endpoint (e.g., LCx, ECx, NOEC, HC5, NOER, NOEAEC, LOEC)
ecotoxvalue	Ecotoxicological value (any decimal number corresponding to the ecotoxicological unit used)
ecotoxunit	Unit of the ecotoxicological value. It could be a soil concentration (mg ai / kg dry weight of soil), or an application rate (kg product/ha)
ecotoxmeasure	Ecotoxicological measure (statendpoint- + measendpoint + duration)

duration	Duration of the test (any decimal number) was used for classification as acute or chronic. It could be any point in time (days, weeks, years) when the measurement was taken, aligned to days after (first) application of the test substance if possible. Not necessary the end of the experimental period.
duration_unit	Unit of the duration
assessarea	Assessment area of the test chemical (e.g., plant, soil macro)
regarea	Registration area of the test chemical (e.g., plant protection products, REACH)

Data were pre-processed, e.g., CAS numbers were aligned to the standard form (hyphenated in the case of the EPA ECOTOX), and additional variables were added so that the origin of the datasets could be traced. Data were filtered to experimental studies dealing with soil-related organism groups as defined above and sent to a PostgreSQL database for storage and performant data retrieval.

The term "endpoint" in ecotoxicology has several meanings. Here, three classes of endpoints, including examples of their usage, were introduced.

- **Measurement endpoints** (*measendpoint*) relate to the biological system on a certain level of organization (e.g., physiology of individuals, characteristics of populations, the structure of communities) and indicate "what was concretely measured by the experimenter". In the laboratory, e.g., mortality, biomass or reproduction are measured, whereas, in the field, mainly abundance, biomass, or biodiversity of several species or groups of soil organisms serve as measurement endpoints.
- The **statistical endpoint** (*statendpoint*) is the result of applying a mathematical model to data, taking decisions on, e.g., significance or quality thresholds according to common sense (e.g., alpha-error probability of five percent in a Williams null-hypothesis test at a power of 80 percent accepted), it indicates "how is toxicity best described".
- The **regulatory endpoint** is of less relevance in the present context because it poses the result of weighing all addressable uncertainties and taking a final decision while a risk assessor analyzes all information available. It indicates "what was decided by the regulator on the acceptable effect level". The latter endpoint was also referred to as the "assessment endpoint" by the Scientific Opinion on the risk assessment of in-soil organisms (EFSA PPR Panel 2017). The regulatory endpoint is not archived in the four data-providing databases and thus not laid down in the *prosoildat* database. An example of a regulatory endpoint would be the RAC – Regulatory Acceptable Concentration

The compilation of data consisted of several individual steps, which build on or complement each other. These steps are organized as a data processing pipeline in individual R scripts, which take over specific tasks in the overall process. The process was designed according to the actual data structures of the source data but independently of the amount and contents thereof so that the final data compilation was fed with new data. In Figure 4 and Figure 5, the single steps and the R scripts are given and briefly explained. A more detailed explanation has been given in the following subchapters and within the original R-scripts (delivered as confidential data accompanying the present report).

Figure 4: Data compilation process structure up to the final data compilation, divided into individual steps that chronologically build upon, or complementing each other. The

individual steps are organized analogously to the data processing pipeline in the individual R scripts, which take over certain specific tasks in the overall process.

00 Load data

00.01_load_data
Loads data from the PROSOIL postgreSQL database & functions into the R-environment

00.02_load_packages
Loads all necessary R packages into R environment

00.03_connect_dbase
Builds up connection to postgreSQL database and ECHA IUCLID Oracle database

00.04_helper_functions
Loads all necessary helper functions into R environment

00.05_helper_convmeasendpoint&filters
Provides conversion strings for measurement endpoints and filter functions

00.07_load_ubacol
Loads UBA corporate design colors into R environment

01 Data import

01.01_import_epaecotox
Import of US EPA Ecotox dat

01.02_import_ubaics
Import of UBA ICS data

01.03_import_ubaetox
Import of UBA ETOX data

01.04_import_echaiuclid
Import of ECHA IUCLID data

01.06_ import_ awident
Import of Alanwoods library data of chemicals

01.08_import_ubaics_ident
Import of the identifier "bearbnr" for products and active ingredients from UBA ICS database

01.09_ import_bvl_ident
Import of registration data and application patterns from BVL

01.10_import_pop_svhc_echacl_ident
Import of ECHA C/L inventory, POP, and SVHC data

01.12_import_chemprops_physprop
Import of physico-chemical parameters from the EPA Physprop database

01.13_import_chemprops_pubchem
Import of physico-chemical parameters from the PubChem database

01.14_import_chemprops_psminfo
Import of physico-chemical parameters from the PSM-Info database

01.15_import_guideline_to_testtype_acute_chr.
Import of harmonization tables for test guidelines to distinguish acute and chronic test types

01.19_import_uba_pbt_list_internal
Import of internal list for assignment of PBT to chemical substances (internal UBA list)

01.23_import_uba_relevant_endpoints
Import of relevant study types assigned by UBA experts manually (expert knowledge)

Source: own illustration, darwin statistics & gaiac

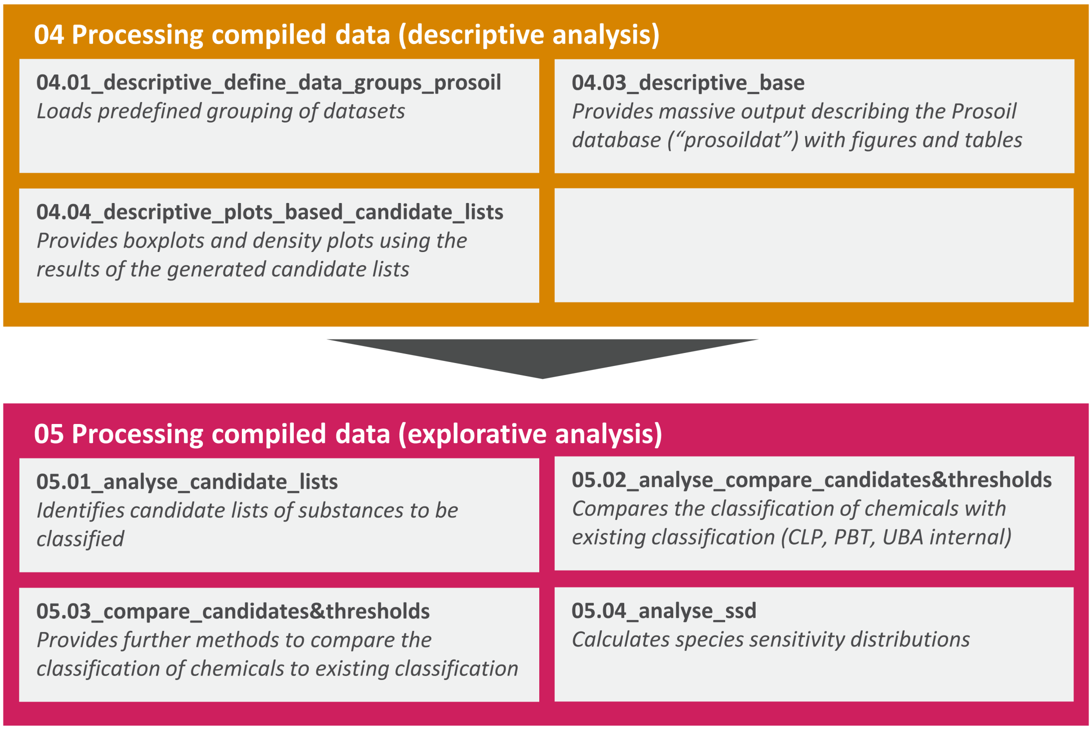
Figure 4 (continued): Data compilation processing structure up to the final data compilation, divided into individual steps, chronologically building upon, or complementing each other. The individual steps are organized analogously to the data processing

pipeline in the individual R scripts, which take over certain specific tasks in the overall process.

Source: own illustration, darwin statistics & gaiac

The definitive data compilation was made available in a project-related PostgreSQL database. All substance-specific data from the different database systems have been linked and stored to provide all available information. All further steps of analysis and grouping were based on this final database. Please note that this database contained protected data and was not publicly available.

Figure 5: Data analysis processing structure starting with the final data compilation divided into individual steps, which chronologically build upon or complement each other. The individual steps are organized analogously to the data processing pipeline in the individual R scripts, which take over certain specific tasks in the overall process.



Source: own illustration, darwin statistics & gaiac

Final data compilation: Two significant milestones were reached after the first two work packages

- The data retrieval was finalized and
- the compilation of data for further analyses in the final project database *prosoildat* was accomplished.

3.1 *Load data* process structure

The subgroup *00 Load data* contains the frame scripts that load all prerequisites (libraries, functions, datasets) necessary for further data processing into the R-Studio environment (IDE - Integrated Development Environment). Several downstream scripts perform different actions, such as establishing a secured connection to the PROSOIL PostgreSQL server to access the stored tables and databases, importing R packages necessary into the R environment, and defining and importing helper functions into the R environment to achieve a proper data processing pipeline. Furthermore, the official UBA corporate design color schemes are loaded into the R environment.

3.2 *Data import* process structure

The subgroup *01 Data import* contains all R scripts that import the original data necessary for further analyses. The scripts include ecotoxicity data, additional identifiers, classification lists, and physicochemical data. The most crucial data import processes are listed in the following.

3.2.1 Import of US-EPA ECOTOX data

This script imports the US-EPA ECOTOX database extract into the PROSOIL PostgreSQL database. The original database extract was downloaded manually from the official US EPA ECOTOX database website (retrieved on 2018-12-13 as 48 separate ASCII files). The data were filtered to soil data by applying the keyword “soil” to the attribute “epa_organism_habitat”, yielding 11.427 entries.

3.2.2 Import of UBA ICS data

This script imports the UBA ICS database extract (previously exported into CSV format from the MS Access database delivered by UBA) into the PROSOIL PostgreSQL database. The ICS data were retrieved on 2020-02-26 as a single long-format table in an MS Access database. The table held 1.282.804 observations (single rows in a table) and 22 variables. The data was then converted to wide-format tables for each of the attributes pre-selected for data export from ICS and considered relevant for the research questions of the PROSOIL project.

Queried ecotoxicity attributes were: "ARTHROP" (mainly non-target arthropods including studies on soil organisms, e.g., collembolans), "MIKRO_TERR" (studies on terrestrial microorganisms, endpoints mainly mineralization rates of artificial substrates), “MULTSP_TER” (terrestrial multispecies tests, field, and semi-field studies, possibly, e.g., terrestrial model ecosystems), “PFLANZEN” (plants, often multiple species from plant protection products efficacy studies), “REGENWURM” (acute and chronic earthworm laboratory studies, also field studies), “SONST_OEK” (other ecological information, no standardized ecotoxicity studies but supplementary information), “SONST_TER (other terrestrial information, similar to “SONST_OEK” from published research data), “SONST_TOX” (other toxicity information, similar the two preceding attributes), “STREU” (litter bag studies).

Queried fate attributes were: “ADS_DESORP” (adsorption and desorption studies), “BIOAK” (bioaccumulation), “BIOAK_TERR” (bioaccumulation from terrestrial exposure scenarios), “BODEN_FREI” (soil degradation field studies), “BODEN_LAB” (soil degradation laboratory studies), “VERF_LUFT” (volatilization into the air).

Queried physicochemical properties attributes were: “DDRUCK" (vapor pressure), “DIS_KON" (dissociation constant), “FETTLOSL” (solubility in fat), “H2OLOSL" (solubility in water), “HEN_KOEF" (Henry coefficient), “LOGPOW” (logarithm of octanol-water partition coefficient), “MOLGEW" (molecular weight), “REL_DICHT” (relative density), “SCHMELZP” (melting point), “SIEDEP” (boiling point), “WIRK_VERW” (substance usage), “WIRKMECH” (mode of action), “ZUS_PC” (additional information on physicochemical properties).

Queried regulatory attributes were: “BEMERKG” (note), “BEW_ANM” (assessment note), “ELINCS" (European List of Notified Chemical Substances), “KENNZG" (labeling), “STOFFBE” (substance description), “ZUSINFO" (additional information).

3.2.3 Import of UBA ETOX data

This script imports an UBA ETOX database extract (.xlsx-format) into the PROSOIL PostgreSQL database. The original database extract was manually queried from both the English and the German version of the system from the website https://webetox.uba.de/webETOX/ as several comma-separated files. The filters in the section “test” were set to “Boden” or “soil”, respectively. Only soil data were imported.

3.2.4 Import of ECHA IUCLID data

This script imports an ECHA IUCLID (Version 6) database extract into the PROSOIL PostgreSQL database (full Oracle database dump amounting to 4.5 GB before re-import). The IUCLID database is accessed via a direct connection to the Oracle database. Only publicly available data, i.e., REACH study results without confidentiality flags, were extracted. The ECHA IUCLID data were retrieved as a dump of the ECHA Oracle database from 2017-02-09, the latest version accessible from the public. No filter criteria were applied, but a full import of REACH study results with 12883 single tables. Subsequently, relevant information and the structure of the database had to be reconstructed from lookup tables holing primary and foreign key fields and from the extensive documentation material. Additionally, REACH study results were also downloaded as an installation file with the desktop version of IUCLID6.

3.3 *Data compiling* process structure

The subgroup *02 Data compiling* contains all R scripts necessary to compile all imported bulk data from *01 Data import* to build the core ecotoxicity datasets using only relevant information. The main steps are the following.

3.3.1 Compiling US-EPA ECOTOX data

This script processes the US-EPA ECOTOX data imported in the previous step. Since the imported records are already filtered data, no further filtering is applied here. The original data entries are assigned to the PROSOIL core variables. The selection of core variables was pivotal in allowing the different database excerpts to be easily merged later. The final step of this script is to export a central data table, ready to be merged with data tables from other sources (*epadat_prefin*).

3.3.2 Compiling UBA ICS data

This script processes the UBA ICS data imported in the previous step. A range of given attributes already filters the imported data; therefore, no further filtering is applied here. Mainly the data tables filtered by the attributes "ARTHROP", "MIKRO_TERR", “MULTSP_TER”, “PFLANZEN”, “REGENWURM”, “SONST_OEK”, “SONST_TER” are further processed. This step is essential for a trouble-free merging with data from the other source in a final data compilation. The final step of this script is to export the data table, ready to be merged with data tables from other sources (*icsdat_prefin*).

3.3.3 Compiling UBA ETOX data

This script processes the UBA ETOX data imported in the previous step. The imported data were already filtered to soil data; therefore, no further filtering is applied here. Within this script, the original entries are assigned to the PROSOIL common database variables to merge with data from the other source in a final data compilation. The final step of this script is to export the data table, ready to be merged with data tables from other sources (*etoxdat_prefin*).

3.3.4 Compiling ECHA IUCLID data

This script processes the imported ECHA IUCLID data tables. The original extract of the IUCLID REACH study results was a complete database dump without any filter. Also, the relations within the data structures had to be rebuilt from scratch. Therefore, the first step here was to search the data for soil datasets only. The imported data table structure is complex, as the datasets are nested and distributed over several tables and linked by primary and secondary keys. The first filter is based on specific PATH variables, defining which compartment or groups of organisms are searched. IUCLID indicates by PATH entries starting with "ENDPOINT_STUDY_RECORD", which information is included in the submitted dossiers. The following data PATH entries have been included in the search.

- "ENDPOINT_STUDY_RECORD.ToxicityToSoilMacroorganismsExceptArthropods"
- "ENDPOINT_STUDY_RECORD.ToxicityToTerrestrialPlants"
- "ENDPOINT_STUDY_RECORD.ToxicityToSoilMicroorganisms"
- "ENDPOINT_STUDY_RECORD.ToxicityToTerrestrialArthropods"
- "ENDPOINT_STUDY_RECORD.ToxicityToOtherAboveGroundOrganisms"

After this filter has been applied, the data belonging to a specific entry are collected by linking primary or secondary keys and combined in a final data table. The final step of this script is to export the data table, ready to be merged with data tables from other sources (*iucdat_prefin*).

3.3.5 Compiling pre-final PROSOIL data

This script merges the final data tables created by the other scripts of this group into a single pre-final table, "prosoildat_prefin", binding all rows of the tables that show precisely the same numbers and column names. This table is then finally exported to the PROSOIL database. Furthermore, secondary functions are applied in these scripts, such as adding row numbers and converting NA information into empty cells.

3.4 Supplementing, harmonizing, and filtering process structure

All R scripts belong here that apply further steps to complement and re-shape the pre-final PROSOIL database ("*prosoildat_prefin*"). These steps complement identifiers, harmonize units and denotations, and remove incomplete data entries. The essential steps are briefly described below.

3.4.1 Add supplementary identifiers

An essential step was to execute a comprehensive algorithm that complemented data entries with missing CAS numbers "aicas" by consulting additional sources and using matches of their common names and IUPAC names.

3.4.2 Harmonizing final ecotoxicity data

The processed data hold a wide variety of different units and levels how the individual processors of the data reported endpoints and test species. The core variables had to be harmonized and converted to gain closed and analyzable datasets and grouped data points. These conversions are essential to ensure the overall comparability of the values and are particularly crucial for identifying and removing duplicate entries from the datasets. Furthermore, the data analysis and grouping strongly depend on harmonized data. Regarding the basic parameters applied in this project, essentially, the variables "*ecotoxunit*", "*testspecies*", "*measendpoint*", "*testtype*", "*duration*", "*durationunit*" and "*statendpoint*" need as much harmonization as possible.

The harmonization process is quite simple in principle: a regular expression (*Regex: a syntactic algorithm that matches strings and thus enables us to develop "find & replace" software procedures*) in the form of an exact match was formulated for every single unique form an entry of a variable can take, regarding the whole database. Expert knowledge was necessary to find and decide the most useful, reliable, and efficient conversion. The main goal of the exercise was to provide as many groups as possible that hold equivalent combinations of study characteristics, i.e., statistical or measurement endpoints and test organisms. The details and rationales of the applied conversions can be found in the appendix.

Some examples illustrate the process of harmonization and conversion.

- "*duration*": Information on study durations has been converted from weeks to days by multiplying the duration by seven, from months to days by a factor of thirty, and from hours to days by dividing the duration by twenty-four.
- "*ecotoxunit*": The best case occurs if the ecotoxunit is already on the same basis so that it can be re-assigned to the desired, unified value, e.g., "mg/kg dry soil" → "mg ai / kg sdw". For example, the situation becomes more complicated for the base ecotoxunit value "kg a.i./ha" that denotes an application rate. The subsequent recalculations can achieve harmonization. The conversion of an application rate given as kg a.i./ha to a soil concentration in a defined soil depth requires two main assumptions.
    - The representative soil density is assumed to be 1.5 g soil/cm^3. The regulatory practice diverges slightly from literature information, e.g., Hartel (2005) mentioned a range of bulk densities from 1.0 to 1.8 g/cm^3 with an average of 1.4 g/cm^3.
    - The appropriate soil depth is set to the upper 5 cm soil layer regardless of substance properties.

3.4.3 Filtering duplicate entries

Duplicate data entries, which found their way into multiple databases, would falsify the shape of distributions of ecotoxicological values (not the ranges or absolute extreme values). Calculation rules are defined that automatically find duplicates. Duplicate entries are particularly likely for databases that focus on collecting studies from the same regulatory context, e.g., UBA ICS and EPA ECOTOX contain studies from experiments with plant protection products. Two rows of data are considered a duplicate if all variables "*aicas*", "*testspecies*", "*ecotoxunit*", "*ecotoxvalue*", "*testtype*", "*statendpoint*", "*measendpoint*", "*duration*", "*durationunit*" have the same value after conversion and harmonization.

3.4.4 Filtering empty entry fields

This script removes observations that have no information in one of the most important core variables "*aicas*", "*testspecies*", "*ecotoxunit*", "*ecotoxvalue*", "*statendpoint*", "*measendpoint*" and "*duration*".

After applying all R scripts for supplementing, harmonizing, and filtering, data retrieval, and compilation were finished. The tables were merged for further analyses in the final PROSOIL project database *prosoildat.*

3.5 Descriptive analysis process structure

The descriptive analysis process structure contains R scripts that mainly apply procedures for overviewing the PROSOIL database ("*prosoildat*") and preparing the data for the derivation of soil toxicity criteria in the follow-up procedural steps of data analysis. The essential scripts are given below.

3.5.1 Selection of predefined criterion groups

This script defines specified PROSOIL criterion groups, namely "*plants acute*", "*plants chronic*", "*soil macroorganisms acute*", "*soil macroorganisms chronic*", and "*soil microorganisms chronic*". The definition of criterion groups is an essential step as all following analyses are based upon the definition of these groups (description of the criterion groups is given in chapter 5).

3.5.2 Basic descriptives

The procedures provide many different descriptive PROSOIL database analyses, which are exported as tables and figures for giving an overview of the contents of the database.

3.5.3 Specific descriptives

These procedures provide specific descriptive PROSOIL database analyses describing the candidate lists generated before (processing in section 3.6, rationales of the derivation of candidate lists are given in section 6.3) via boxplots and probability density plots.

3.6 Explorative process structure

The explorative algorithms represent the core of developing soil toxicity criteria for CLP and PBT assessments that were applied to the final PROSOIL database ("prosoildat"), piecewise for each of five criterion groups (section 5). The main scripts are described below.

3.6.1 Identify candidate lists

This procedure is the primary step toward the PROSOIL candidate lists. It identifies all substances included in a specific approach's candidate lists. Hence, the resulting list contains all three threshold values that classify substances as "*very toxic*", "*toxic,*" or "*harmful*" in a specific statistical approach.

3.6.2 Comparison with existing classification systems

This script compares the PROSOIL candidate lists with the ECHA C&L inventory and PBT status classifications. Therefore, all substances within a specific candidate list are checked for their status in the existing classification. The comparison result, i.e., the status, is finally added to the candidate list.

3.6.3 Species sensitivity distributions

This script analyses the species sensitivity distributions within the PROSOIL candidate lists and exports the distributions as figures and tables.

4 Descriptive analysis of the *prosoildat* database

The *descriptive data analysis* of the final data compilation was considered essential to facilitate the reader's access to an assessment of data reliability and was embedded at various stages of the data compilation process. The analyses were performed (1) to describe the status of data compilation before the explorative analysis, (2) to provide insight into the data quality and quantity, and (3) to transparently present the before-after course of the data that has undergone numerous steps of harmonization and grouping between the different data sources and of filtering out to assure the minimum data quality standards.

The database "*prosoildat*" used to define data groups for data analysis and the development of soil toxicity criteria comprised more than 125.000 data entries and 66 variables. It covered the results of experimental studies from five regulatory areas of chemicals, namely the REACH regulation, the regulations for biocides, human and veterinary medicinal products, and plant protection products. The database was complemented by numerous datasets from the peer-reviewed scientific literature. In particular, the scientific literature covered the diversity of possible test organisms with more than 2.700 distinct species. The variety of organic chemicals and metals reflected the whole history of ecotoxicological research. The database contained more than 3.700 different chemicals, as indicated by the unique identifier CAS-number. The "*prosoildat*" posed a solid database further described in the following.

4.1 Mandatory information, quality, and completeness checks

A set of core variables was defined as the most essential and primary criterion for data usability (i.e., in the context of developing soil toxicity criteria by applying exploratory statistical data analysis). The set of core variables was considered **mandatory** to hold valid information. It consisted of the following variables (in bold letters appear the variable names as present in the database):

- ***aicas*** The CAS registry number was taken as the unique identifier across the entire data analysis workflow. It provides the most consistent identifier of a chemical and, even if it is far from perfect, contains fewer ambiguities than, e.g., the IUPAC name and was already included in most international data repositories used in the PROSOIL project.
- ***ecotoxvalue*** A valid data entry must hold a concrete numerical value. Without this value, no quantitative analysis was possible, and rows containing empty entries were discarded.
- ***ecotoxunit*** Each ecotoxicological value had to be accompanied by a specific unit, at best soil concentrations or application rates, but also more exotic units were seen as noted by the original database processors.
- ***statendpoint*** The statistical endpoint is as essential as the ecotoxicological value, and without it, no grouping of comparable datasets for the derivation of soil toxicity thresholds could be made.
- ***duration*** The soil toxicity criteria were derived using datasets grouped by temporal characteristics of the underlying experiments into "short-term" or "acute and "long-term" or "chronic" studies. If this assignment was not given initially, it had to be deduced from the duration or the provisions of the standardized guidelines used. Here, the understanding of the duration did not relate to the whole period between the experiment's start and termination. The term duration was considered the point in time after the experiment's start when the measurement or observation was taken. Classifying a study as "acute" or "chronic" is sometimes difficult, however, for many tests defined via the test protocol. Well-established

definitions were taken from the guidance in EPS 1//RM/45 (Environment Canada 2005). It states that "acute" means measuring effects within a short exposure period (seconds, minutes, hours, or a few days) and not concerning the organism's life span.

- In contrast, the term "chronic" refers to effects that occur within a relatively long exposure period (weeks, months, or years), usually comprising a significant portion of the organism's life span, such as ten percent or more, and includes the observation of reproduction success. There were several cases where the complete sequence and history of an experiment were documented in the database. In such a case, the "pedantic" scientific approach, the assignment of "acute" or "chronic" could have changed for a specific study with increasing test duration. On the other hand, the "pragmatic" approach reported the regulatory relevant endpoint at the end of the study only, and there was an unambiguous assignment of acute or chronic possible.
- ***durationunit*** As with the ecotoxicological value-unit pair, the duration must be given in a standard unit of time (so it is accessible to conversions in principle). It was aimed at deriving the duration in days without exception.
- ***testspecies*** A valid entry of the tested organismic entity must be given. The entity could be a taxonomic species or genus name in Latin or common language notation or a larger group like a "soil community" for field testing.

Only datasets with complete entries in the set of mandatory variables were included in the exploratory data analysis to derive soil toxicity criteria. The overall effect of removing data from the last pre-final version of the *prosoildat*-database after removing duplicate entries (refer to section 4.2) can be reviewed by inspecting the corresponding variables in Table 4, which reflects the status before filtering the datasets for the core variables.

Table 4: Numbers of missing information for seven core variables in the *prosoildat*-database. The last prefinal version 4 of the database contained 162.509 entries in total.

variable	number	percent
aicas	8441	5.19
ecotoxvalue	4807	2.96
ecotoxunit	4842	2.98
duration	18757	11.54
durationunit	6930	4.26
statendpoint	6755	4.16
testspecies	6616	4.07

The overall least information with 12 % data entries missing was given for the variable *duration*. The fact that the (descriptive) units of *duration* were more often provided than the (numeric) duration is due to the practice that particularly for plant data periods were given, e.g., "to bloom", "to emergence", or "from larva to adult".

After completing the data harmonization, various quality checks were carried out to identify errors. These quality checks consisted of a manual plausibility check of the individual database entries. The most critical quality checks were the following:

- very low *ecotoxvalue* entries were checked for plausibility and correct conversion during the harmonization process (which was closely related to the harmonization of *ecotoxunit*). The threshold value for this quality check was *ecotoxvalue* < 10^{-6}.
- very high *ecotoxvalue* entries were also checked for plausibility and correct conversion during the harmonization process (i.e., harmonization of *ecotoxunit*). The threshold value for this quality check was *ecotoxvalue* > 10^{6}.
- *duration* entries were checked for plausibility and correct conversion during the harmonization process (i.e., harmonization of *durationunit*). The threshold values for this quality check were 0.5 d < *duration* < 365 d.

In the event of irregularities during the harmonization or data analysis, individual database entries were repeatedly checked and corrected if necessary. Such a detailed quality check was not possible for all database entries.

Data availability is highly regulation-specific due to different data requirements and other historical factors that differ from chemical legislation. Soil toxicity data for products under the REACH regulation were scarce, whereas for compounds falling under the regulation of plant protection products, soil data belonged to the core information requirements. However, the regulatory origin of the data was not primarily crucial at this point but rather as a general quality check for completeness of the final data compilation in "*prosoildat*" that was carried out. A completeness analysis of missing entries across all variables can be seen in Table annex 1. The number of missing entries per variable and registration area was quantified for this deficit analysis.

The original databases UBA ICS, UBA ETOX, EPA ECOTOX, or ECHA IUCLID were queried across the variables for complementary information on missing data, e.g., chemical identifiers or physicochemical properties. Additional data sources were matched against the identifiers of the existing data tables. As can be seen from Table annex 1, the most significant number of missing entries occurred for compounds classified as literature data (*regarea* LITRA) and REACH data (*regarea* REACH). However, most missing values were observed for the variable class "physicochemical properties" and some chemical identifiers, which do not significantly impact the data analyses. Thus, a case-wise consideration of missing entries was conducted after grouping in the specific groups and analyzing subsets of the data (including ad-hoc handling of missing values for computing averages or connected statistical characteristics).

4.2 Harmonization and duplicate removal

The harmonization of the data entries consisted of two processes:

1. **Renaming** of character values to a uniform term
2. **Converting** numerical values to the base of a uniform unit.

Several examples of how these two mechanisms impacted the nature of the data are given in Table 5. After the harmonization process, the structure of the data changed clearly. Since no data was deleted from the database, the number of entries remained the same. **Duplicate entries** were only removed after harmonization (for details and sequence of the processing algorithms, refer to chapter 3). Three characteristic states during the processing of the database showed the profound effects of the implemented changes.

- State 1: The data structure directly after the essential database compilation from UBA ICS, ECHA IUCLID, US EPA ECOTOX, UBA ETOX as "*prosoildat_prefin*"
- State 2: The data structure after all harmonization steps
- State 3: The data structure after the duplicate removal (the criteria were described in chapter 3), i.e., after the final data compilation was completed.

All harmonization steps and the duplicate removal were necessary to make the single data entries comparable to each other and allow for the comparative analysis. Table 5 shows the most common conversions of the essential core variables.

The **renaming** process was done manually and necessitated expert knowledge but no re-calculation of the ecotoxicological values.

Data harmonization considerably impacted the original data structure from the **conversion** of the ecotoxicological units (Figure 6), and data harmonization considerably impacted the original data structure. This step converted the *many different units* of the ecotoxicological values used within the four data-providing databases UBA ICS, ECHA IUCLID, UBA ETOX, and EPA ECOTOX by countless human data processors over decades of ecotoxicological history *into a small number of unified units*. Whenever possible from expert judgment and while knowing the experimental studies' underlying concepts, the original unit was converted to the most frequent and widely accepted unit, "soil concentration", given as "mg ai/kg sdw" (mg active ingredient per kg soil dry weight). However, conversion was impossible in each case and strongly depended on the test type.

The conversion process required several specifications: 1.) an algorithm for how the ecotoxicological value would be changed when the original unit was transposed to the desired unified unit. The algorithm consisted of multiplication and division steps of the original units using conversion factors. 2.) the appropriate conversion factors had to be defined, as shown by Table 5, e.g. the conversion of kg active ingredients per hectare (AI kg /ha) to mg ai / kg sdw needed the conversion factor from kg ai to mg ai [*kg.to.mg.factor], the hectare to square-meter factor [/ha.cm2.factor] and the soil weight at a given standard density of 1.5 at a given standard soil depth of 5 cm [/soil.weight.at.depth.factor].

The reference values of the ecotoxicological units were often familiar to European regulatory frameworks because international databases from the Anglo-Saxon area included certain conversion factors, e.g., pound (lbs) to kg or acre to square-meter were necessary.

Table 5: Harmonization procedure for exemplary variables (VAR) according to the conversion step's frequency (i.e., only the conversions with the highest frequency FREQ are shown).

VAR	FREQ	BEFORE	AFTER	CONV
testspecies	187	Senecio vulgaris, Poa annua, Andryala integrifolia, Hypocharis radicata	Senecio vulgaris & Poa annua & Andryala integrifolia & Hypocharis radicata	renaming
	136	Zea mays, Solanum tuberosum, Nematoda	Zea mays & Solanum tuberosum & Nematoda	renaming
	102	Folsomia candida, I. viridis	Folsomia candida & Isotoma viridis	renaming
	72	Cenococcum geophilum, exposed in Fries solution	Cenococcum geophilum	renaming
	32	Terrestrial plants, QSAR model	Plantae	renaming
	23	A. caliginosa, E. crypticus	Aporrectodea caliginosa & Enchytraeus crypticus	renaming
	19	other terrestrial plants: lettuce, canola, and barley	Lactuca sativa & Brassica rapa & Hordeum vulgare	renaming
ecotoxunit	48230	mg/kg soil dw	mg ai/kg sdw	renaming
	24759	AI kg/ha	mg ai/kg sdw	*kg.to.mg.factor/ha.cm2.factor/soil.weight.at.depth.factor
	18285	kg/ha	mg ai/kg sdw	*kg.to.mg.factor/ha.cm2.factor/soil.weight.at.depth.factor
	14671	AI lb/acre	mg ai/kg sdw	*lbs.to.mg.factor/acre.to.hectar.factor/ha.cm2.factor/soil.weight.at.depth.factor
	8647	mg/kg	mg ai/kg sdw	renaming
	7865	AI g/ha	mg ai/kg sdw	*g.to.mg.factor/ha.cm2.factor/soil.weight.at.depth.factor
	6545	g/ha	mg ai/kg sdw	*g.to.mg.factor/ha.cm2.factor/soil.weight.at.depth.factor
	5684	mg/L	mg ai/l	renaming
statendpoint	64252	NOEL	NOEC	renaming
	9466	LC_{50}	EC_{50}	renaming
	5290	ER_{50}	EC_{50}	renaming
	4236	LD_{50}	EC_{50}	renaming
	3236	LR_{50}	EC_{50}	renaming
	1313	NOER	NOEC	renaming
	1175	IC_{50}	EC_{50}	renaming
measendpoint	10132	mortalität	Mortality	renaming
	9990	mortality	Mortality	renaming
	5016	reproduktion	Reproduction	renaming

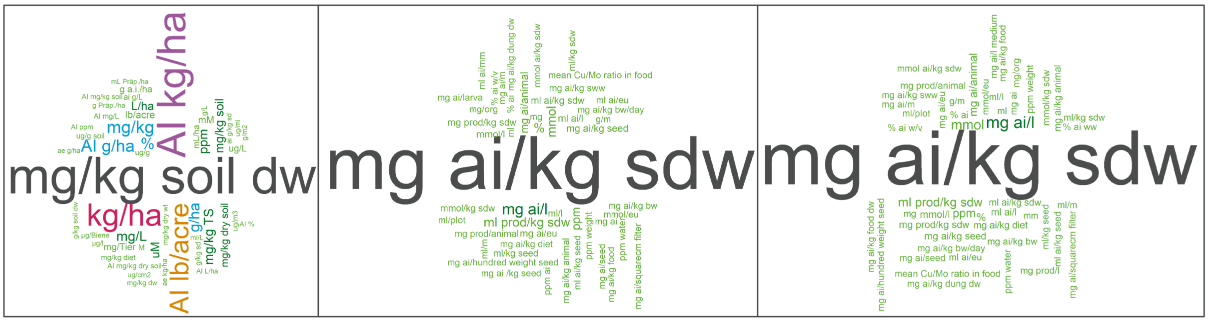
VAR	FREQ	BEFORE	AFTER	CONV
	2328	nitrate formation rate	nitrate_formation_rate	renaming
	2315	seedling emergence	seedling_emergence	renaming
	1902	respiration rate	respiration_rate	renaming
	1828	nitratkonzentration	nitrate_concentration	renaming
	1708	gewicht	Weight	renaming
	1618	frischgewicht	fresh_weight	renaming
	1596	vigor	vegetative_vigour	renaming
	1594	co2-entwicklung	carbon_dioxide_formation	renaming
	1437	root elongation	root_elongation	renaming

Conversions CONV for the variables *testspecies*, *ecotoxunit*, *statendpoint*, and *measendpoint*. The conversion factors for *ecotoxunits* were defined in separate R-scripts and lookup tables, and the conversion was implemented as a sequence of multiplication and division steps on the original *ecotoxvalues*.

A **duplicated data entry** was defined as holding identical information in at least nine variables of the ore dataset: "*aicas*", "*testspecies*", "*ecotoxunit*", "*ecotoxvalue*", "*testtype*", "*statendpoint*", "*measendpoint*", "*duration*", "*durationunit*". If such a combination was detected for two or more data entries, regardless of the data-providing database where the entry was originally stored, the first entry was kept, and the following entries were discarded. The process was transparent by adding a new variable, "is_duplicate", to the prefinal data table before removing the duplicate entries. Since several variables ordered the data table alphabetically, preceding entries (e.g., with the *origin* EPA) were kept while subsequent entries (e.g., the origin ICS) were discarded. The method led to several entries of the same study integrated by more than one data-providing database being deleted. The mechanism worked particularly well if the duplicate was recognized after harmonization and could not be detected before.

Most of the harmonized variables consisted of many different data entries, making a detailed visualization of all conversions impossible. Figure 6 shows three exemplary representations as "word clouds", showing the relative share in most frequent terms used as ecotoxicological units was stored in the variable "*ecotoxunit*" before the harmonization steps and duplicate removal (left), after harmonization (center), and after duplicate removal (right). The word clouds aided in the visualization of the process; all individual entries were made available in Excel workbooks.

Figure 6: Word clouds showing the evolution of distribution and assignment of ecotoxicological units (the variable *ecotoxunit*) from state 1 before harmonization and duplicate removal steps (left) to the state 2 distribution after all harmonization steps (center) and the final state 3 *prosoildat* database after duplicate removal (right).



Source: own illustration, darwin statistics & gaiac

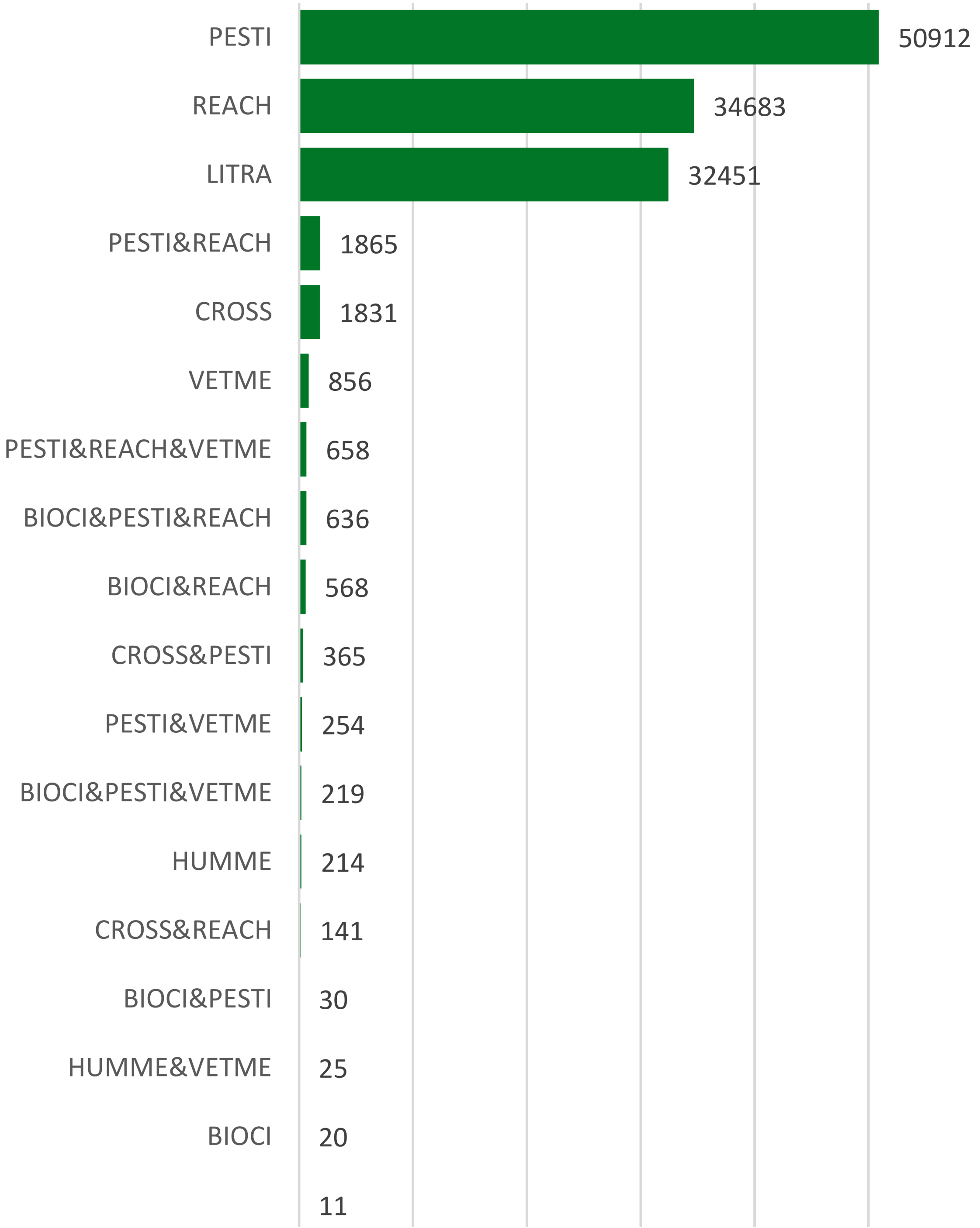
The distribution and frequencies of the statistical endpoints (*statendpoint*) entries were impacted by both the data harmonization (e.g., due to unit conversion) and the duplicate removal. During the first step, many different endpoint notations (e.g., NOEC vs. NOEL or EC_{50} vs. ER_{50}) were converted into standardized formats. The rationale was that the measurement endpoint, e.g., “mortality” or “lethality”, indicated in combination with the *ecotoxunit* the circumstances under which the *statendpoint* was determined sufficiently and must not be signaled by the *statendpoint*

As demonstrated for the variable *ecotoxunit* (Figure 4), all core variables were analyzed for relative share and distribution changes before and after the **harmonization** and **duplicate removal**. The assignments were adapted and optimized repeatedly in an iterative process.

The duplicate removal step had only a minor impact on the " ecotoxunits " data structure because the number of removed data entries was relative and impacted similar proportions of entries across the original units.

The *testspecies* data were harmonized by assigning uniform Latin species names (so they could be matched with taxonomic data from GBIF-repositories) while eliminating different spelling and name variants. The duplicate removal step had only a minor impact on the relative occurrence frequency and, in this case, only led to a heterogeneous background distribution, whereas the main test species remained prevalent. The highest efforts to reach the best possible harmonization of the *testspecies* variable did not appear at the most frequent entries. It was more evident at the very long tail of less frequent but very heterogeneous possible entries of test species notations, to the point that numerous tested species were combined into one single data entry. From 3.663 different *testspecies* identifiers, 2.718 entries with proper taxonomic nomenclature (species names composed as taxonomic binomials from generic and specific epithets or higher taxonomic orders if not further specified as the term “*soil community*” or “*plantae*”) remained after conversion in the final *prosoildat* database.

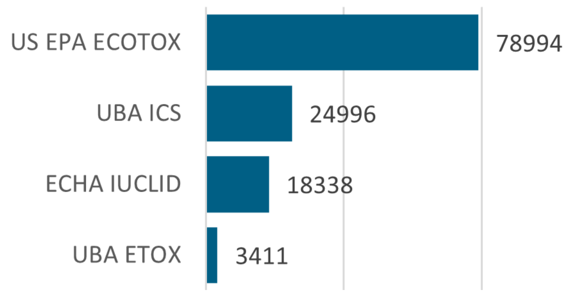
Figure 7: Regulatory context of the data compilation.



Source: own illustration, darwin statistics & gaiac

Numbers refer to the number of entries in the database, not to unique studies. PESTI = plant protection products (PPP), some authors use the term "pesticides" equivalently to PPP, in the context of European legislation frameworks, it could include biocides and PPP; REACH = chemicals regulated under REACH; LITRA = information from publications from scientific literature, no regulatory area assignable; CROSS = substance regulated by more than one unique regulatory framework, definition made available from UBA ICS, also indicated by the &-operator; VETME = veterinary medicinal products; HUMME = human veterinary medicinal products; BIOCI = biocidal products.

Figure 8: Origins of the data compilation.



Source: own illustration, darwin statistics & gaiac

Numbers refer to the number of entries in the database, not to unique studies.

4.3 Origin of the data and assignment of the regulatory context

After harmonizing data, the first step was assigning the *origin* (the data-providing database in the variable *origin*) of the data and the context under which the substance was regulated (variable *regarea*) based on the metadata therein. The databases contained information from studies generated by notifying or registering companies and publicly available academic studies. For chemical data originating from peer-reviewed literature that could not be assigned by cross-reading within the US EPA database to a regulatory area, the respective entry was flagged as "LITRA" in the metadata *regarea* in the database. Information for this analysis was directly obtained from the data compilation as the origin was defined by the data source (ECHA IUCLID, UBA ICS, US EPA, UBA ETOX).

The area of regulation, e.g., REACH chemicals, plant protection products, biocides, or human or veterinary pharmaceuticals, defined the regulatory context (Figure 7). Both analyses were closely linked, and some of the databases covered predominantly specific regulatory contexts (e.g., ECHA IUCLID = REACH chemicals or biocidal active substances). The definitions and assignments to regulatory areas mainly followed the classification provided by the UBA ICS database (the original attribute "ART" which translates to "type" in the English language, was utilized). Even though the legislation under which a substance has to be registered and regulated depends primarily on the use pattern of the respective substances and substances identified via their CAS numbers, scientific publications do not refer to specific legal provisions. Attempts were made to assign a substance's most probable regulatory background by looking up the *prosoildat* database for matches with the same substance indicating the regulatory area. The substance was categorized as "literature data" (LITRA) when derived from the EPA ECOTOX and UBA ETOX databases and was not assignable to any regulatory areas despite lookup.

Most of the available soil data came from the pesticides regulation (PESTI), followed by REACH chemicals and literature sources (LITRA) with an unknown regulatory background. Comparatively few data on biocides (BIOCI) effects were available, frequently found regulated under various legislation. The coverage of veterinary and human medicinal products (HUMME and VETME) assignments was scarce and originated from the UBA ICS database.

Most of the REACH data (REACH) were taken from the ECHA IUCLID database, and most plant protection products data (PESTI) data were derived from the UBA ICS database. It had to be kept in mind that certain substances were assigned to more than one regulatory context, as the specific substance revealed multiple use patterns, and the test data were consequently used under more than one legislation. The latter resulted in combined assignments of the regulatory areas, i.e., REACH, plant protection products, biocides, veterinary substances, and pharmaceuticals (see Figure 7). If a database entry was assigned to LITRA, the source was the open, peer-reviewed literature. If it was set to CROSS, multiple regulatory areas were considered by the UBA ICS database (without detailed information available on which exact areas were covered). The distribution of the origins of the data was inserted in Figure 8.

4.4 Distribution of ecotoxicity between test organisms from different regulatory assessment areas

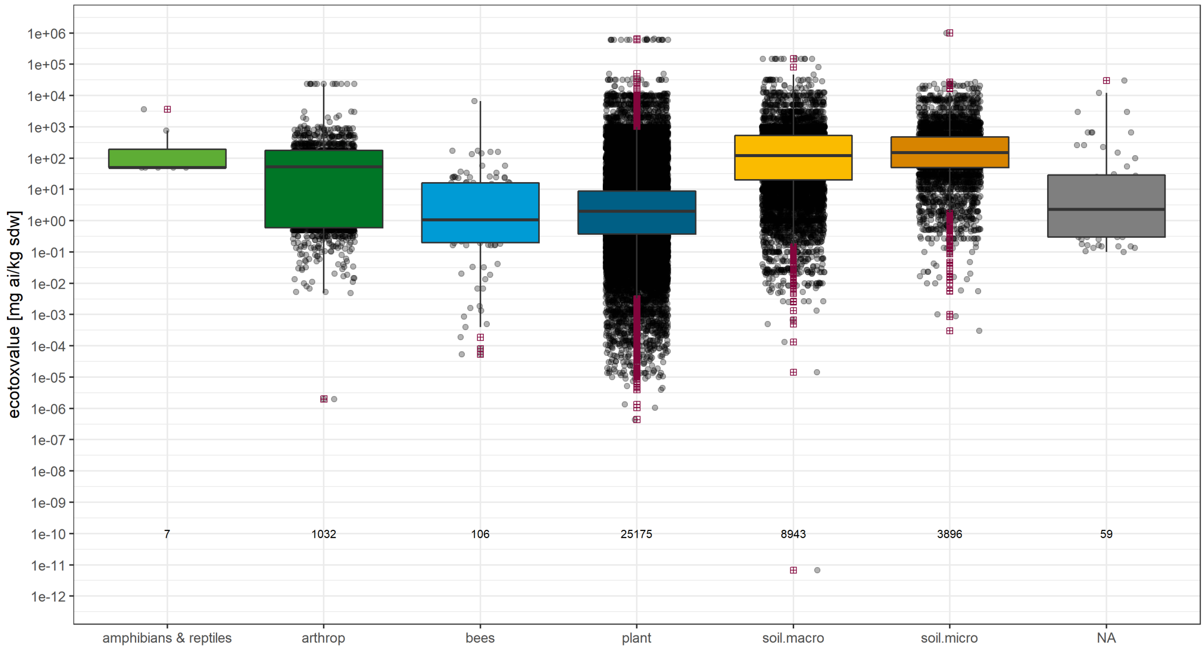
The chemical legislation that was the focus of the present work were biocides, plant protection products, REACH chemicals, and medicinal products (human and veterinary). Even though substantial differences in data requirements exist between the regulatory areas, the environmental risk assessment schemes have one thing in common: they are organized in largely independent risk assessment areas (variable *assessarea* in the *prosoildat* database). The classification used was leaned against the German scheme (derived from the UBA ICS database that filed the assessment area within the attribute "MERKMAL") that relied on the European system for its part. The assessment areas can be divided into aquatic and terrestrial environmental compartments. The latter could be further distinguished between above-ground and below-ground assessments. Terrestrial assessment areas classify into birds and mammals, non-target plants (category *plant* in *prosoildat*), non-target arthropods (category *arthrop*), bees (category *bee*), in-soil macroorganisms (category *soil-macro in prosoildat*, attribute "REGENWURM" and "ARTHROP" filtered for soil organisms from UBA ICS database), and in-soil microorganisms (category *soil-micro in prosoildat*). No specific risk assessment scheme for amphibians and reptiles is available. However, if incidences of endocrine effects appear, data on amphibians is also required within the evaluation of PPP. Initially, the database was compiled **using data from all terrestrial assessment areas available** from the four data-providing databases.

Having the toxicity database complete and at hand, initial questions on the overall shape of distributions of ecotoxicological values were central. The overview served as decision support for which groups should be further used to derive toxicity thresholds and hazard classifiers, interacting with the introductory provisions of the aims and scope of the present research project (chapter 1.6). It was central for the PROSOIL approach that the focus was strictly on valid in-soil organisms as defined in chapter 1.4.

Were there organism groups representative of certain risk assessment areas of the various regulatory contexts (e.g., plant protection products or REACH regulations)? Was it possible to identify organism groups that appeared to hold lower ecotoxicity values than others, which could mean they would be more sensitive towards the chemicals notified in general, or they were in favor of the data requirements under the most dominant regulatory regimen? Which groups were the most data-rich over all datasets compiled? Figure 9 below shows the distribution of all terrestrial assessment areas before filtering on actual in-soil organisms, as defined in chapter 1.4. In principle, it turned out to be necessary to include as many groups as possible in the first instance. Many data processors classified the studies with non-target arthropods as arthropods from a taxonomic point of view (e.g., *Hypoaspis aculeifer*) as being essential components of the in-soil organism standard test battery (for plant protection products). After the initial data import, the opportunity was to re-assign the assessment areas.

The database held a massive variety of different levels of the core variables data entries. More than 120 distinct ecotoxicological units and more than 100 statistical endpoint categories were reported, revealing considerable differences in representation, being the *ecotoxunit* "mg ai /kg sdw" and the *statendpoint* "NOEC" the most frequent entries. Consequently, the minimum comparability between any data points of a consistent analysis should be ensured by filtering for common statistical endpoints and ecotoxicological units. Here, it was filtered to the "NOEC" (often but not necessarily used for chronic toxicity values) and the ecotoxicological unit "mg ai / kg sdw". It has to be kept in mind that specific assessment areas use their standard studies, resulting in specific exposure regimens and thus ecotoxicological units. For example, the standard studies for bees expose the individuals by dermal or oral application of the test item, and thus the ecotoxunit is given as the "amount of active ingredient per individual" and could not be converted into a soil concentration. Many data are given as application rates and could be accordingly converted into soil concentrations as it was common regulatory practice for the risk assessment of in-soil organisms. The representativeness of Figure 9 must be perceived with these general conditions in mind.

Figure 9: Distribution of ecotoxicological values of six assessment areas covered by the *prosoildat* database



Source: own illustration, darwin statistics & gaiac

Additionally, data of unknown assessment areas "not available" = NA; ecotoxunit = "mg ai / kg sdw", statistical endpoint = NOEC. Outlying points (1.5 IQR) in dark fuchsia color.

The boxplots showed that the distributions of ecotoxicity values of most "risk assessment areas" spanned a wide range of about ten orders of magnitude, with still multiple values on both the higher and the lower end of the distributions (the outliers were identified by the 1.5-fold interquartile range-method and marked in dark fuchsia color in Figure 9). Formally outliers, these values reflected the normal range of ecotoxicity data, considering the wide span between different organism groups and test systems (sometimes embracing several decades of ecotoxicological research and regulatory practices) and covering various physicochemical properties of the tested chemicals. Ecotoxicological studies using plants as test organisms were the most frequent category (dark blue box in Figure 9) and showed a very high absolute number of values beyond the interquartile range. These "outliers" covered between five (*soil-macro*) and seventeen percent of all data points (*plants*). The analysis overall assessment areas could be interpreted as plants were frequently tested in a test battery of various cultivated and wild species while applying a wide range of application rates (here converted from rates to soil concentrations using standard assumptions).

The median values for plants were lower than soil macro and microorganisms. The lower median was probably because the conversion from surface application rates (i.e., kg ai / ha) to soil concentrations (i.e., mg/kg sdw) was not valid for the organisms since exposure was not realized via soil. In follow-up analyses for plants as proper in-soil organisms, a relevance flag based on the concrete test system (refer to the chapter "Criterion groups for soil organisms" for the derivation of filter criteria that helped to derive consistently comparable groups) served to filter exposure scenarios where the plants were not treated directly via soil.

The focus of the present work was the derivation of hazard criteria for in-soil organisms. Experiments using terrestrial organisms were conducted under various exposure regimens that were not considered well comparable. The filtering to a common set of ecotoxicological units and statistical endpoints demonstrated that only very few data for bees (although a standard requirement for PPP), amphibians, and reptiles were available. For these reasons and from the above analysis, it was decided that only data from the groups of plants (if the life-stage was exposed via soil), soil macroorganisms, and soil microorganisms were included in the further analyses. Additionally, community and field studies were excluded from the consistent *criterion groups* (for definitions, see chapter 5).

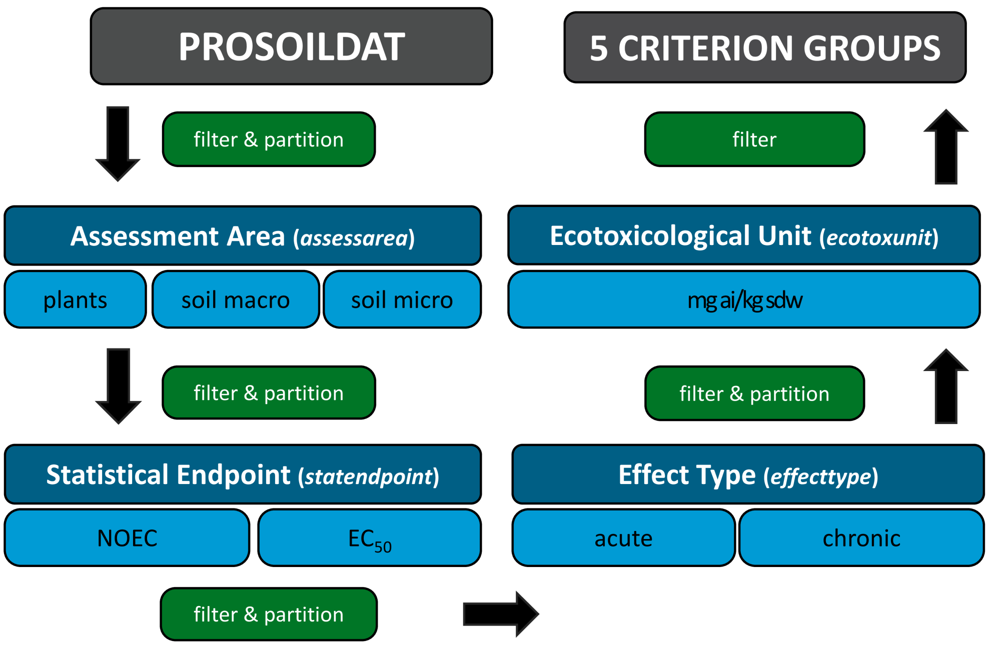
5 Criterion groups for soil organisms

5.1 Evaluation of criterion groups for soil organisms

The database held a great diversity of measurement- and statistical endpoints from various groups of test organisms. Before a detailed data analysis could be conducted, the data had to be filtered to a homogenous set of data entries. First and foremost, the grouping and filtering of the complete *prosoildat* database into coherent and comparable data groups was essential to deriving soil toxicity criteria. These groups were defined and used as **criterion groups** in all further chapters. Together with the data harmonization (see chapter 4.2), the primary aim of the data grouping was to make the data comparable to each other through classification (i.e., "not to compare apples and oranges"). During the harmonization, data were converted to reduce the number of different units and notations in favor of few but well comparable test systems representing typical organism groups and exposure situations. This step was essential as a comparison of data between different categories, such as ecotoxicological units (e.g., soil concentrations vs. application rates) or statistical endpoints (e.g., results of null-hypothesis significance testing vs. linear regression), or from different risk assessment areas (e.g., experiments with higher plants vs. soil organisms) was considered non-acceptable and scientifically unsubstantiated.

The grouping of datasets into criterion groups obtained in this way was thus considered the fundamental basis for the following analyses, e.g., to deduce soil toxicity criteria in CLP and PBT assessments. For a better understanding of the grouping method, the hierarchical decision tree is shown in Figure 10.

Figure 10: Procedural steps of building five criterion groups for the separate development of soil toxicity thresholds by a decision tree. The definitions of the variables used for filtering were explained in section 3 and complemented by section 4.2.



Source: own illustration, darwin statistics & gaiac

5.1.1 First step - risk assessment area (*assessarea*)

The data were first divided into three groups based on the assessment areas *plants* by filtering the entry *"plant"*, microorganisms using the entry "soil micro," and soil macroorganisms with the entry "soil macro" (all criteria were listed by Table 6). This filtering and partitioning step defined the organism groups that were focused on as true in-soil organisms and set the scene for the further definition of criterion groups (as can be seen from the decision tree in Figure 10). This step excluded birds & mammals, amphibians & reptiles, and non-target arthropods, including bees, from the datasets that were finally analyzed. The other mentioned risk assessment areas could be subjected to broader analyses of terrestrial organisms in follow-up activities.

5.1.2 Second step - statistical endpoint (*statendpoint*)

The second step of the grouping procedure subdivided the three groups from the risk assessment areas into the most frequent statistical endpoints. Within the *prosoildat* database, the NOEC was predominantly frequent, providing more than 40 % of all entries (53.000 entries out of 125.000 rows). The EC_{10} was considered equivalent to the NOEC. However, this *statendpoint* was given in 7 % of the cases only. Both statistical endpoints were used for the chronic study results. NOEC values for acute toxicity tests were excluded if they did not belong to a chronic study design according to the German Environment Agency UBA.

The EC_{50} was seen in about 20 % of the cases in the *prosoildat* database (25.000 entries) and was mainly used for acute measurement endpoints. The statistical endpoints of "soil macroorganisms with the *ecotoxunit* mg ai/kg sdw" were homogeneously distributed between NOEC (2.844 entries) and EC_{50} (2.485 entries). The statistical endpoints originating from "soil microorganisms" with the *ecotoxunit* "mg ai/kg sdw" showed 286 entries with NOEC or EC_{10}. Finally, plants revealed 2.462 observations for the *statendpoint* EC_{50} and 7338 observations for the NOEC.

5.1.3 Third step - effect type (*effecttype*)

Plants and soil organisms were further divided into "acute" and "chronic" studies using the variable *effecttype*, while all experiments using soil microorganisms were assumed to measure chronic endpoints only.

5.1.4 Fourth step - ecotoxicological unit (*ecotoxunit*)

Many different ecotoxicological units were already converted to the key unit of this study, i.e., "mg ai/kg sdw". The criterion groups were filtered to unique entries of this unit, which was solely used as the commonly agreed most relevant unit. Due to a harmonization step (chapter 3.4.2), the study results for the assessment area "*plants*" were to a great extent used with the *ecotoxunit* "mg ai / kg sdw". For this unit, approx. 57.000 data entries were available. The same applied to the assessment area "*soil macroorganisms*". Here, the data mainly consisted of the ecotoxicological unit "mg ai / kg sdw" with approx. 23.000 data entries. Last, the assessment area "*soil microorganisms*" consisted of approx. 7.400 data entries for the ecotoxicological unit "mg ai / kg sdw". Hence, the remainder of 27.5 % of the 125.000 entries with other ecotoxicological units were not used for further analyses.

5.1.5 Additional step 1 - expert judgment

Further expert judgment with the knowledge of the specific standard test procedures and used endpoints was applied in some instances to check the database for correct assignments, e.g., in cases where a NOEC is used for an acute study or to assign the valid statistical endpoints to plant studies exposed via soil.

5.1.6 Additional step 2 - cut-off values (*operator*)

Only studies with true NOECs were considered and studies with the attribute "*operator*" equalled ">", "<", ">=", "<=", "ca.", "~" were excluded. A NOEC as the No Observed Effect Concentration is defined only if at least one test concentration has shown significant effects, and a Lowest Observed Effect Concentration LOEC can be derived. This filter criterion excluded a considerably high number of data entries (around 13 % of total *prosoildat*). Another possibility would have been to proceed similar to Renaud et al. (2004) to neglect the LC_{50}s "smaller than" but set the LC_{50} ">" or ">=" as "=".

They wrote: *"When a ">" operator preceded an LC_{50} value, the value following the operator was used (e.g., LC_{50} = 1000 if the data was LC_{50} > 1000 mg kg^{-1})."*

Due to the limitations of the standard regulatory test procedures (e.g., limit testing), NOECs can often not be defined precisely. The cut-off of specific ecotoxicological values, be it "lower than" or "greater than" a true, in this case, non-measured and not proven threshold value (i.e., NOEC), comes with certain costs. The distribution of ecotoxicological values was censored at the highest tested concentration, often fixed to a smoothed number, e.g., 1000 or 2000 mg ai /kg sdw.

In this context, it must be considered that Renaud et al. dealt with LC_{50}-values derived by regression analysis (e.g., Probit or 3-parameter logistic models could have been used) from acute earthworm toxicity studies after 14 days of exposure. The information inherent in a dose-response design is more valuable than a limit test design and including the censored values could be more reliable because a clear effect level could be derived from inspecting the dose-response curve. It remained questionable whether a proper, regulatory valid and statistically sounded dose-response design would not deliver a determinable LC_{50}-value.

On the other hand, if several larger and smaller values were added to the ends of the distribution of ecotoxicological values, likewise as Renaud et al. did, it would not change its shape. However, it would become broader, i.e., the measure of variation like the standard deviation would become greater while the measures of central tendencies like the arithmetic mean would not necessarily.

Here, we cut off the larger values only. This procedure resulted in a distribution of ecotoxicological values that could be shifted or skewed towards smaller *ecotoxvalues*. The methods to deduce toxicity thresholds (described in chapter 6) that rely on the order of ecotoxicological values would be based on a shorter list of potentially classified substances and lower threshold values. The lower the threshold value, the fewer substances would be classified. In theory, cutting the upper end of the distribution could be less protective. Renaud et al. stated from their analyses that more than 70 % of the data contained values with the ">"-operator, and they used two separate datasets to derive cut-off values, one with and one without censored data on the upper end of the distribution. The result was a cut-off value for the reduced database of 2 mg ai / kg sdw that was approximately 50 % lower than the entire database (4 mg ai / kg sdw).

The selection of filter specifications that fix the criterion groups for the definition of soil toxicity criteria (as outlined in chapter 6), the grouping was dependent on the number of data sets per criterion group and the requirements of a hazard-based assessment of chemicals. If this number is low, no reliable and generalizable explorative data analysis could be performed. However, soil microorganisms were included, although the data set was small, holding 286 data entries only. Table 6 shows the five criterion groups used to derive soil toxicity criteria and the resulting data entries within each criterion group.

5.2 Selection criteria and number of data entries per criterion group

After homogeneous groups of comparable data have been formed, the resulting group size, i.e., the number of datasets per criterion group, is crucial for the reliability of the following (statistical) data analysis. On the other hand, the grouping was limited to a few steps to achieve a compromise between data comparability on the one hand and a sufficient group size.

Table 6: Characteristics of five criterion groups.

criterion group	assessarea	ecotoxunit	statendpoint	effecttype	n data entries
group1	plant	mg ai / kg sdw	EC_{50}	Acute	2462
group2	plant	mg ai / kg sdw	NOEC + EC_{10}	Chronic	7338
group3	soil macro	mg ai / kg sdw	EC_{50}	Acute	2485
group4	soil macro	mg ai / kg sdw	NOEC + EC_{10}	Chronic	2844
group5	soil micro	mg ai / kg sdw	NOEC + EC_{10}	Chronic	286

5.3 Toxicity distribution of criterion groups

The five above pre-defined criterion groups were assessed regarding their concrete shape of distributions of ecotoxicity values. This analysis was done to answer whether the five criterion groups were homogenous enough to be assumed to belong to a single, common random sample as a prerequisite of subsequent statistical analyses. It was possible to conduct an expert elicitation of the distributions after clear definitions of risk assessment areas related to **true in-soil organisms,** transparent descriptions of the **criterion groups,** and clear definitions of the **quality criteria**. Consistent toxicity criteria from uniform procedures depend on toxicity distributions of the criterion groups that do not systematically differ intrinsically, i.e., certain organism groups (e.g., mites) within the criterion group (e.g., soil macroorganisms) of the data should not be more sensitive than others (e.g., earthworms). In the case of soil macroorganisms, the dataset was compiled of eight taxonomic subgroups, namely *collembolans*, *dungorganism, enchytraeids, isopods, lumbricids, mites, myriapods, and nematodes*, which differ in the history of evolving ecotoxicological test methods, trophic positions in soil food-webs and their importance in current soil risk assessment procedures.

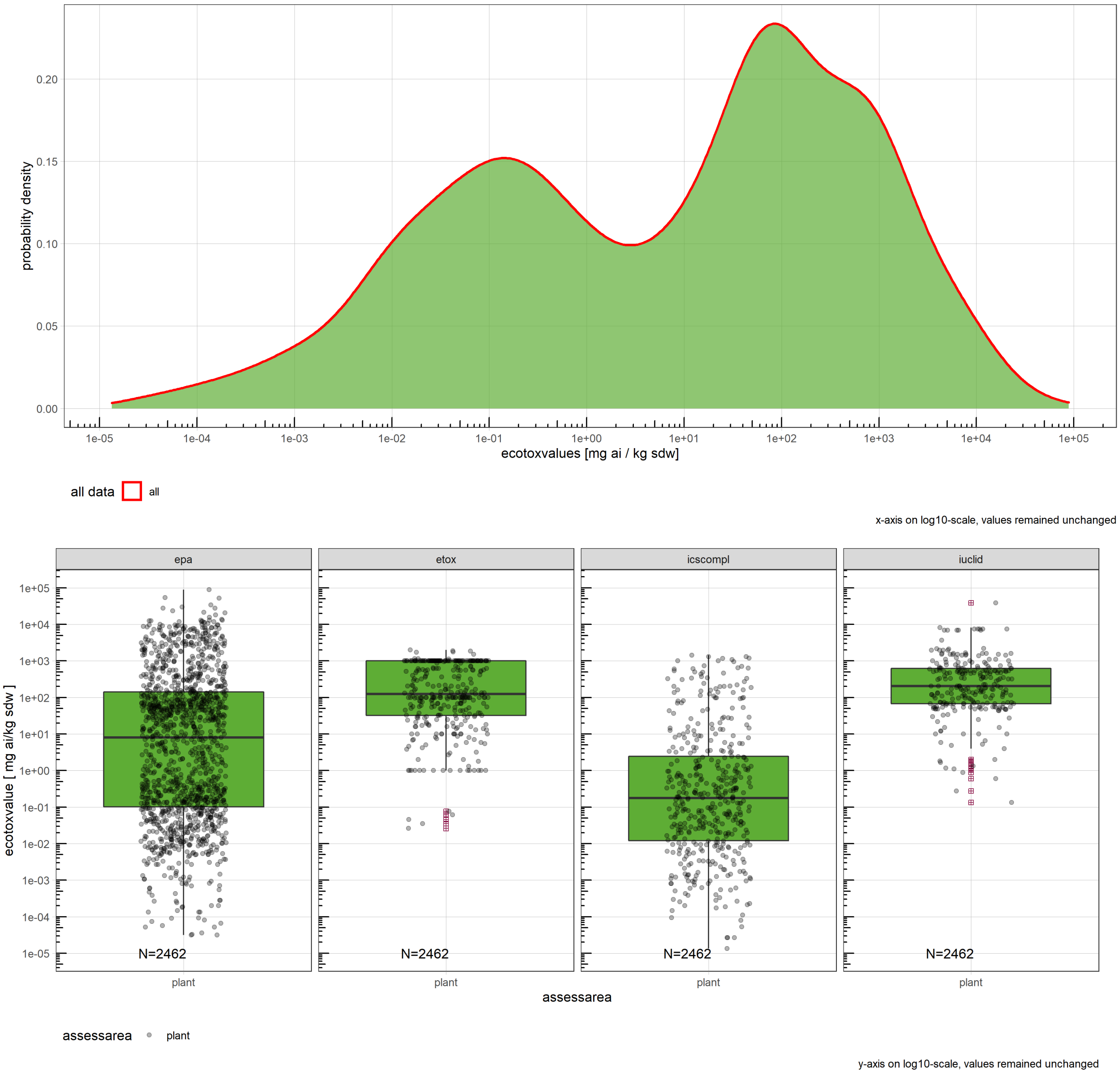
The interpretations for the inclusion of the five criterion groups are presented in the following. Frequency distributions of ecotoxicological values and data characteristics were revisited thoroughly, and the rationales why these particular groups were to be included are explained (detailed tables and figures can be found in the supplementary material).

One remark to evaluating the data on toxicity to plants, i.e., the usability of the criterion groups "plants-acute" and "plants-chronic," might be hampered by the database entry system of the UBA ICS database. The UBA ICS included only data of the most sensitive species out of six to ten tested species. Most likely, this procedure was not applied by the processors of three other databases, which, therefore, much better reflected the range of sensitivities of plants and should be considered better suited for scientific questions and not for regulatory purposes only.

5.3.1 Distribution of plants acute data – criterion group 1

The criterion group "plants acute" toxicity entries covered a wide range of ecotoxicological values from 10 ng ai / kg sdw to 10.000 mg ai / kg sdw. The 2.462 data observations for the criterion group 1 "*plants acute*" were highly dominated by more than 50 % of the data with the origin (ORIG) of the US EPA database.

Figure 11: Distribution of soil ecotoxicity values from acute studies with plants (criterion group 1).



Source: own illustration, darwin statistics & gaiac

The density plot shows all values, and the boxplots show the distributions split into the data-supplying repositories.

Table 7: Statistical characteristics of the ecotoxicological values of criterion group 1.

ORIG	MEAN	MEDIAN	GEO	MIN	MAX	LQ10	LU10	N
epa	974.79	8.05	4.54	<0.01	88527.20	0.01	2480.00	1402
etox	409.71	123.50	112.39	0.03	2000.00	5.80	1000.00	352
ics	53.14	0.18	0.21	<0.01	1433.33	<0.01	108.80	448
iuclid	860.84	204.00	195.16	0.13	38621.00	28.50	1558.00	260

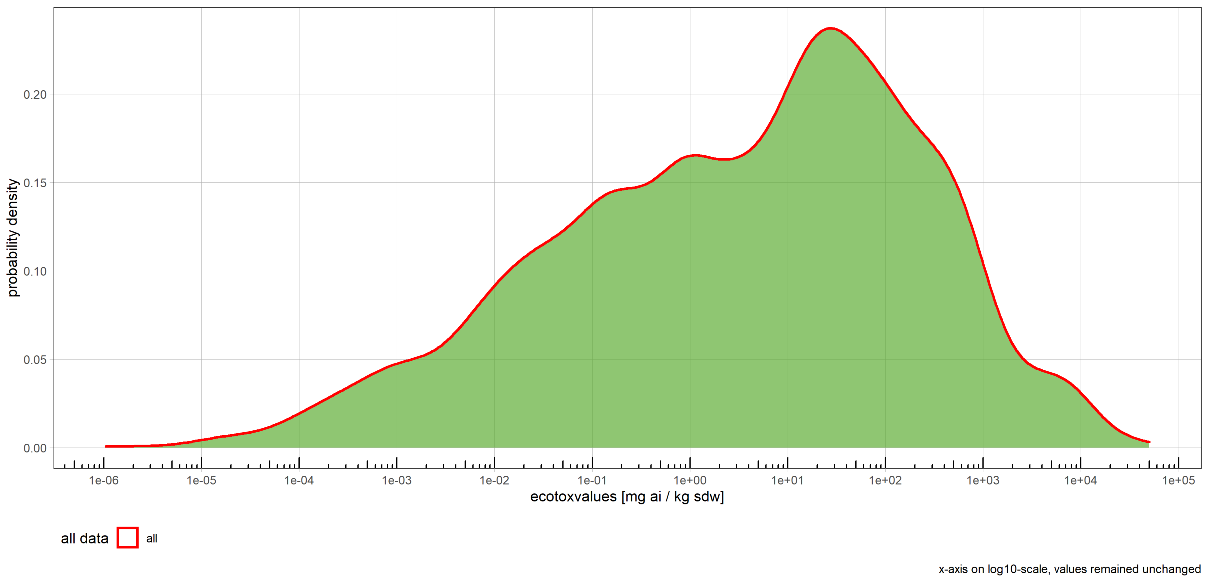
ORIG: the origin of the data as the data-providing repository (US EPA ECOTOX, UBA ETOX, IBA ICS, ECHA IUCLID); MEAN: arithmetic average of the ecotoxicological values of the criterion group; MEDIAN: median of the ecotoxicological values of the criterion group; GEO: the geometric mean of the ecotoxicological values of the criterion group; MIN: minimum of the ecotoxicological values of the criterion group; MAX: maximum of the ecotoxicological values of the criterion group; LQ10: lower 10th quantile of the ecotoxicological values of the criterion group; LU10: upper 10th quantile of the ecotoxicological values of the criterion group; N: sample size as the number of entries in the criterion group; all numbers except the sample size are given as mg ai / kg sdw

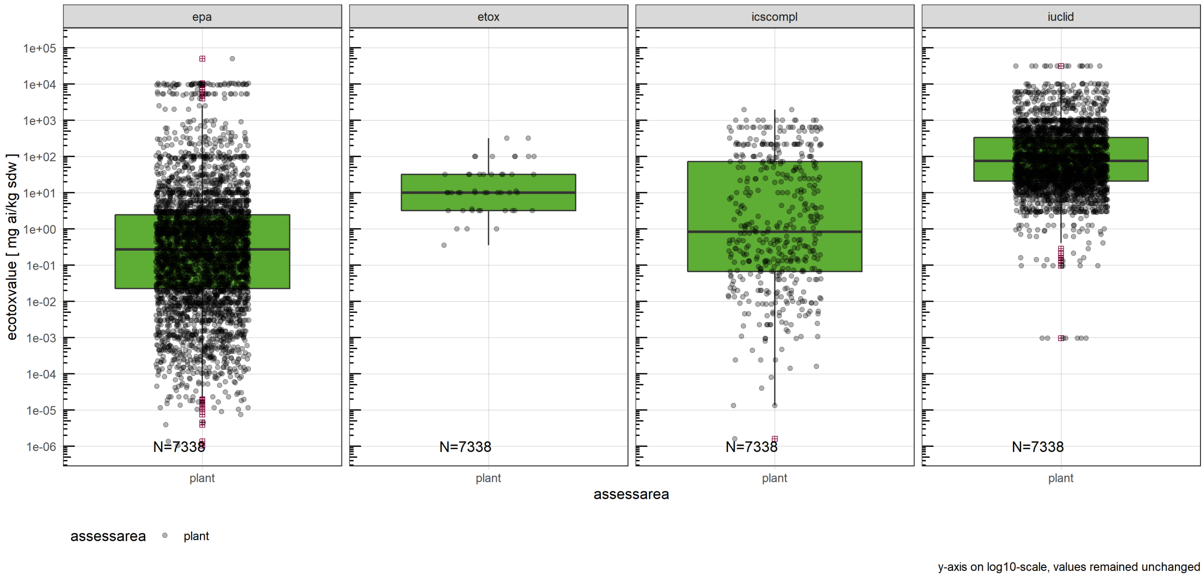
Two distinct distribution peaks were identified, at around 100 mg ai / kg sdw and at 0.1 mg ai / kg sdw (Figure 11). UBA ICS data delivered lower toxicity values, ascribed to the plant protection product risk assessment data requirements, predominantly reflected by the ICS data. On the other hand, the different databases came from industrial chemicals (IUCLID) or peer-reviewed publications (EPA, ETOX) and tended to cover the whole diversity of toxicities much better than the ICS database (Table 7).

5.3.2 Distribution of p*lant chronic* data – criterion group 2

The criterion group "*plants chronic*" toxicity entries covered ecotoxicological values ranging from 1 µg ai / kg sdw to 50.000 mg ai / kg sdw. The 7.338 data observations for the criterion group 2 "*plants chronic*" were highly dominated by more than 50 % of EPA ECOTOX datasets. The distribution peak was around 400 mg ai / kg sdw (Figure 12).

Figure 12: Distribution of soil ecotoxicity values from chronic studies with plants (criterion group 2).





Source: own illustration, darwin statistics & gaiac

The density plot shows all values, and the boxplots show the distributions split into the data-supplying repositories.

UBA ICS data delivered lower toxicity values (mean value of 102 mg ai / kg sdw), ascribed to the plant protection product risk assessment data requirements, predominantly reflected by the ICS data. On the other hand, the different databases came from industrial chemicals (very high average vales for ECHA IUCLID data of 630 mg ai / kg sdw) or peer-reviewed publications (EPA, ETOX). They tended to cover the diversity of toxicities much better than the ICS database.

Table 8: Statistical characteristics of the ecotoxicological values of criterion group 2.

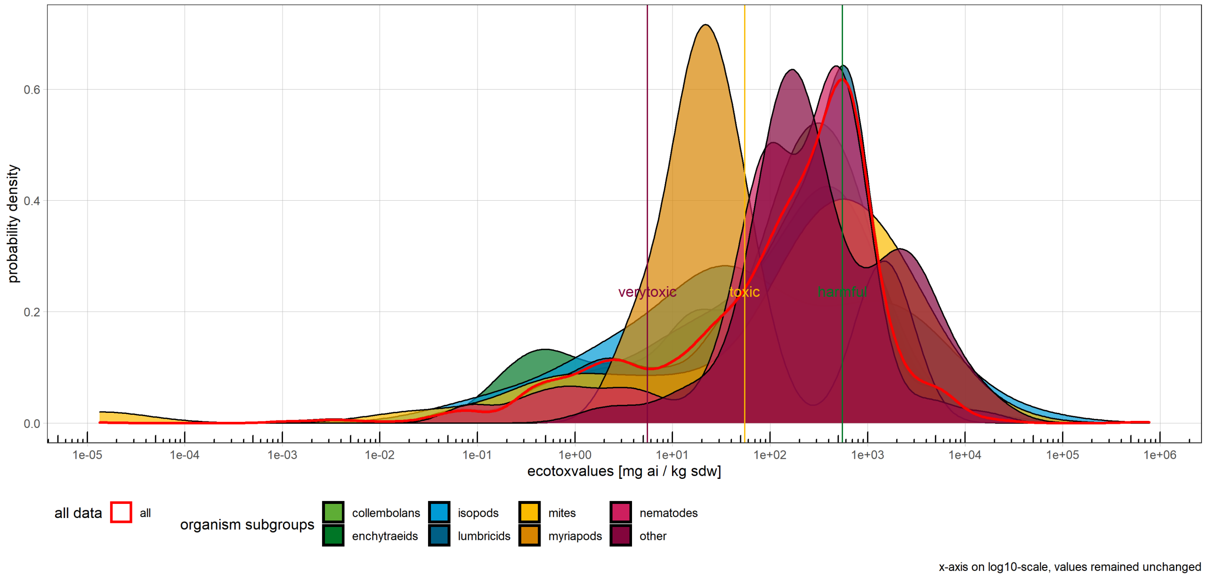
ORIG	MEAN	MEDIAN	GEO	MIN	MAX	LQ10	LU10	N
epa	206.34	0.27	0.24	<0.01	50000.00	<0.01	20.67	3797
etox	39.10	10.00	13.27	0.35	320.00	3.20	100.00	47
ics	102.13	0.83	1.26	<0.01	1960.00	0.01	303.00	454
iuclid	630.47	74.90	82.96	<0.01	31428.00	8.60	1000.00	3040

ORIG: the origin of the data as the data-providing repository (US EPA ECOTOX, UBA ETOX, IBA ICS, ECHA IUCLID); MEAN: arithmetic average of the ecotoxicological values of the criterion group; MEDIAN: median of the ecotoxicological values of the criterion group; GEO: the geometric mean of the ecotoxicological values of the criterion group; MIN: minimum of the ecotoxicological values of the criterion group; MAX: maximum of the ecotoxicological values of the criterion group; LQ10: lower 10th quantile of the ecotoxicological values of the criterion group; LU10: upper 10th quantile of the ecotoxicological values of the criterion group; N: sample size as the number of entries in the criterion group; all numbers except the sample size are given as mg ai / kg sdw

5.3.3 Distribution of *soil-macro acute* data – criterion group 3

The criterion group "soil-macro acute" toxicity entries covered a wide range of ecotoxicological values from 10 µg / kg sdw to 770.000 mg ai / kg sdw.

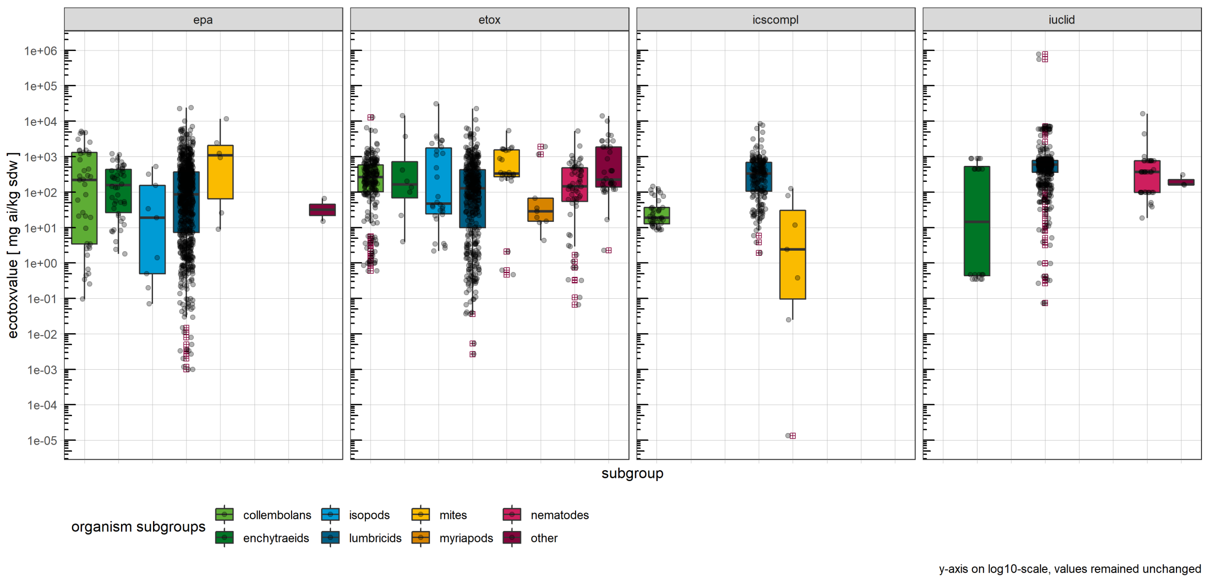
Figure 13: Distribution of soil ecotoxicity values from acute studies with soil-macroorganisms (criterion group 3).



Source: own illustration, darwin statistics & gaiac

The density plot shows all values, and the boxplots show the distributions split into the data-supplying repositories and all subgroups.

Figure 13 (continued): Distribution of soil ecotoxicity values from acute studies with soil-macroorganisms (criterion group 3). The density plot shows all values, and the boxplots show the distributions split into the data-supplying repositories and all subgroups.



Source: own illustration, darwin statistics & gaiac

The density plot shows all values, and the boxplots show the distributions split into the data-supplying repositories and all subgroups.

Table 9: Statistical characteristics of the ecotoxicological values of criterion group 3.

SUB	ORIG	MEAN	MEDIAN	GEO	MIN	MAX	LQ10	LU10	N
collembolans	epa	914.95	217.00	69.99	0.10	5150.00	0.50	3100.00	41
collembolans	etox	510.18	265.00	167.95	0.60	12784.00	5.26	917.00	225
collembolans	icscompl	36.31	18.70	24.65	8.50	145.00	10.00	95.40	41
enchytraeids	epa	281.89	156.50	102.00	1.80	1200.00	7.90	799.00	46
enchytraeids	etox	2338.63	168.00	218.51	4.00	14150.00	4.00	14150.00	8
enchytraeids	iuclid	334.21	223.24	16.00	0.35	890.00	0.35	890.00	20
isopods	epa	116.90	18.80	8.16	0.07	523.00	0.07	523.00	9
isopods	etox	1845.92	47.70	132.31	2.19	31000.00	3.25	3091.50	30
lumbricids	epa	460.48	86.04	46.57	<0.01	24195.00	0.90	1083.00	759
lumbricids	etox	467.13	128.50	58.15	<0.01	22371.00	0.76	1014.00	446
lumbricids	icscompl	558.62	331.50	241.37	1.93	8473.00	27.30	920.00	196
lumbricids	iuclid	3600.35	594.50	423.74	0.07	770000.00	97.00	865.00	476
mites	epa	2726.00	1089.00	445.51	9.00	11689.00	9.00	11689.00	6
mites	etox	882.94	333.00	285.73	0.47	5406.00	2.10	1720.00	25
mites	icscompl	31.51	2.40	0.62	<0.01	126.20	0.00	126.20	7
myriapods	etox	360.23	28.60	51.43	4.29	1910.00	4.29	1910.00	9
nematodes	etox	370.95	146.00	81.99	0.07	5300.00	1.10	710.00	62
nematodes	iuclid	921.97	365.00	289.50	18.60	16042.80	50.20	768.00	36
other	epa	40.50	40.50	31.46	15.00	66.00	15.00	66.00	2
other	etox	1501.31	223.00	409.44	2.28	13900.00	100.00	3980.00	38
other	iuclid	208.58	164.82	199.09	158.13	302.79	158.13	302.79	3

SUB: taxonomic subgroup of soil-macroorganisms; ORIG: the origin of the data as the data-providing repository (US EPA ECOTOX, UBA ETOX, IBA ICS, ECHA IUCLID); MEAN: arithmetic average of the ecotoxicological values of the criterion group; MEDIAN: median of the ecotoxicological values of the criterion group; GEO: the geometric mean of the ecotoxicological values of the criterion group; MIN: minimum of the ecotoxicological values of the criterion group; MAX: maximum of the ecotoxicological values of the criterion group; LQ10: lower 10th quantile of the ecotoxicological values of the criterion group; LU10: upper 10th quantile of the ecotoxicological values of the criterion group; N: sample size as the number of entries in the criterion group; all numbers except the sample size are given as mg ai / kg sdw

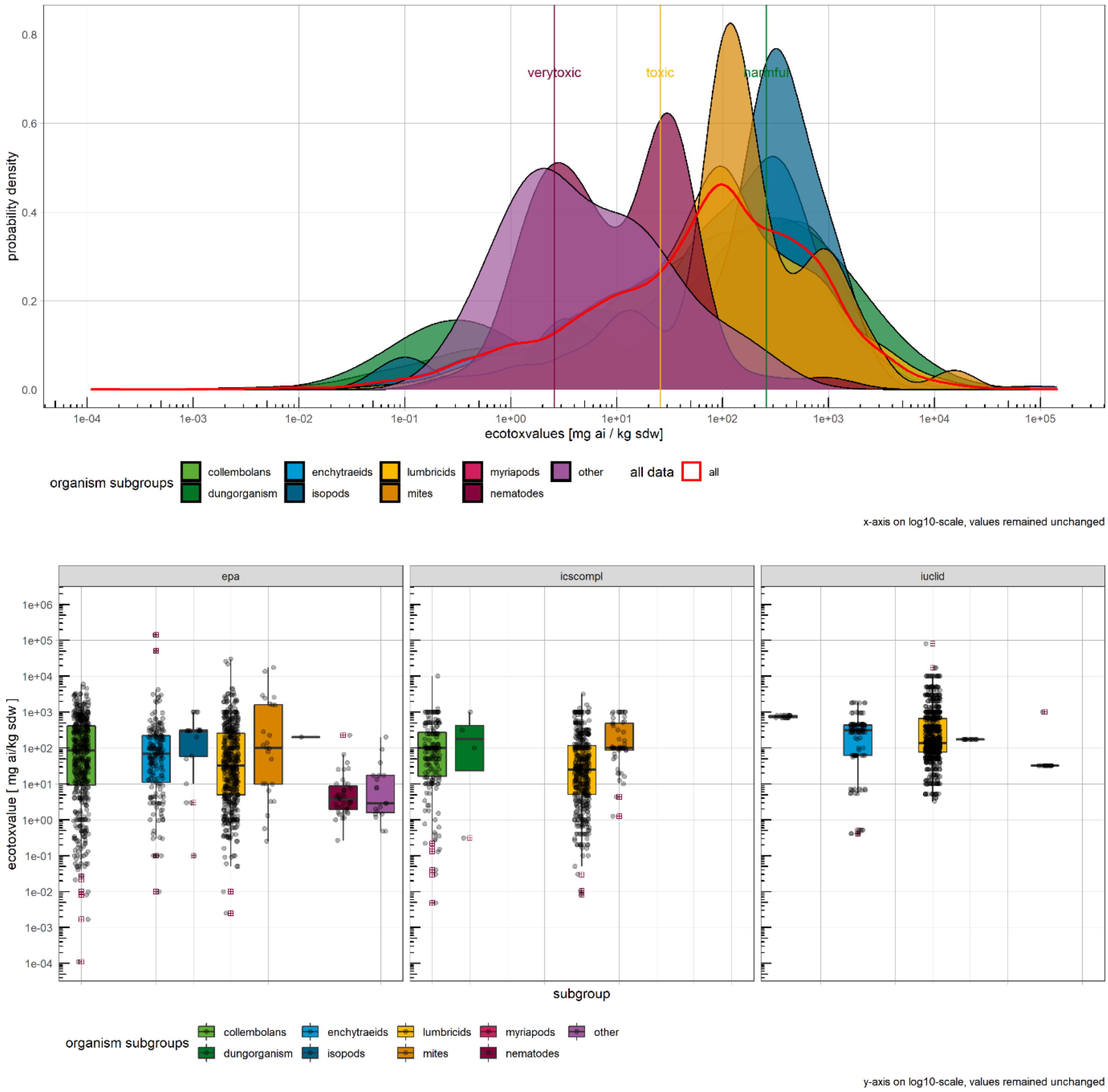
The 2.485 data observations for criterion group 3 - "*soil-macro acute*" were not highly dominated by one data-providing database. Each species subgroup (SUB) showed distinct distribution peaks, from collembolan data at 0.3 mg ai / kg sdw to enchytraeids at 400 mg ai / kg sdw (Figure 13). UBA ICS data delivered lower toxicity values, ascribed to the plant protection product risk assessment data requirements, predominantly reflected by the ICS data.

On the other hand, the different databases came from industrial chemicals (IUCLID) or peer-reviewed publications (EPA, ETOX) and tended to cover the whole diversity of toxicities much better than the ICS database (Table 9).

5.3.4 Distribution of soil-macro chronic data – criterion group 4

The data origin of the datasets for soil macro-organisms (chronic studies) was more equally distributed over the data suppliers US EPA, UBA ICS, and ECHA IUCLID.

Figure 14: Distribution of soil ecotoxicity values from chronic studies with soil-macroorganisms (criterion group 4).



Source: own illustration, darwin statistics & gaiac

The density plot shows all values, and the boxplots show the distributions split into the data-supplying repositories and all subgroups.

The UBA ETOX database contained only study types assigned “acute” and is therefore not shown in figure 14. In conclusion, the differences from splitting the ecotoxicity distributions between the eight taxonomic subgroups *collembolans, dungorganism, enchytraeids, isopods, lumbricids, mites, myriapods, and nematodes* (and *others*) were considered not relevant.

Table 10: Statistical characteristics of the ecotoxicological values of criterion group 4.

SUB	ORIG	MEAN	MEDIAN	GEO	MIN	MAX	LQ10	LU10	N
collembolans	epa	367.98	83.80	43.10	0.00	6000.00	0.48	987.00	469
collembolans	icscompl	277.68	100.00	57.38	0.00	10000.00	2.50	1000.00	179
collembolans	iuclid	750.00	750.00	748.33	700.00	800.00	700.00	800.00	20
dungorganism	icscompl	354.08	208.00	55.95	0.31	1000.00	0.31	1000.00	4
enchytraeids	epa	1296.58	68.25	46.37	0.01	142356.00	1.90	500.00	184
enchytraeids	iuclid	381.65	312.00	133.14	0.41	1889.00	5.44	947.00	114
isopods	epa	352.91	300.00	113.14	0.10	1000.00	3.00	1000.00	20
lumbricids	epa	419.69	32.00	31.14	0.00	30000.00	1.00	800.00	497
lumbricids	icscompl	141.43	25.00	21.90	0.01	3200.00	0.89	445.00	436
lumbricids	iuclid	914.70	136.20	197.98	3.30	80800.00	37.50	2404.00	750
mites	epa	1798.55	100.00	90.54	0.25	17496.00	1.30	2896.00	25
mites	icscompl	300.52	100.00	129.31	1.25	1000.00	19.20	1000.00	54
mites	iuclid	174.00	174.00	174.00	174.00	174.00	174.00	174.00	13
myriapods	epa	200.00	200.00	200.00	200.00	200.00	200.00	200.00	1
nematodes	epa	13.87	3.73	4.53	0.27	227.00	1.12	26.67	39
nematodes	iuclid	80.40	32.00	38.01	32.00	1000.00	32.00	32.00	20
other	epa	22.49	2.88	5.48	0.48	200.00	0.48	90.67	19

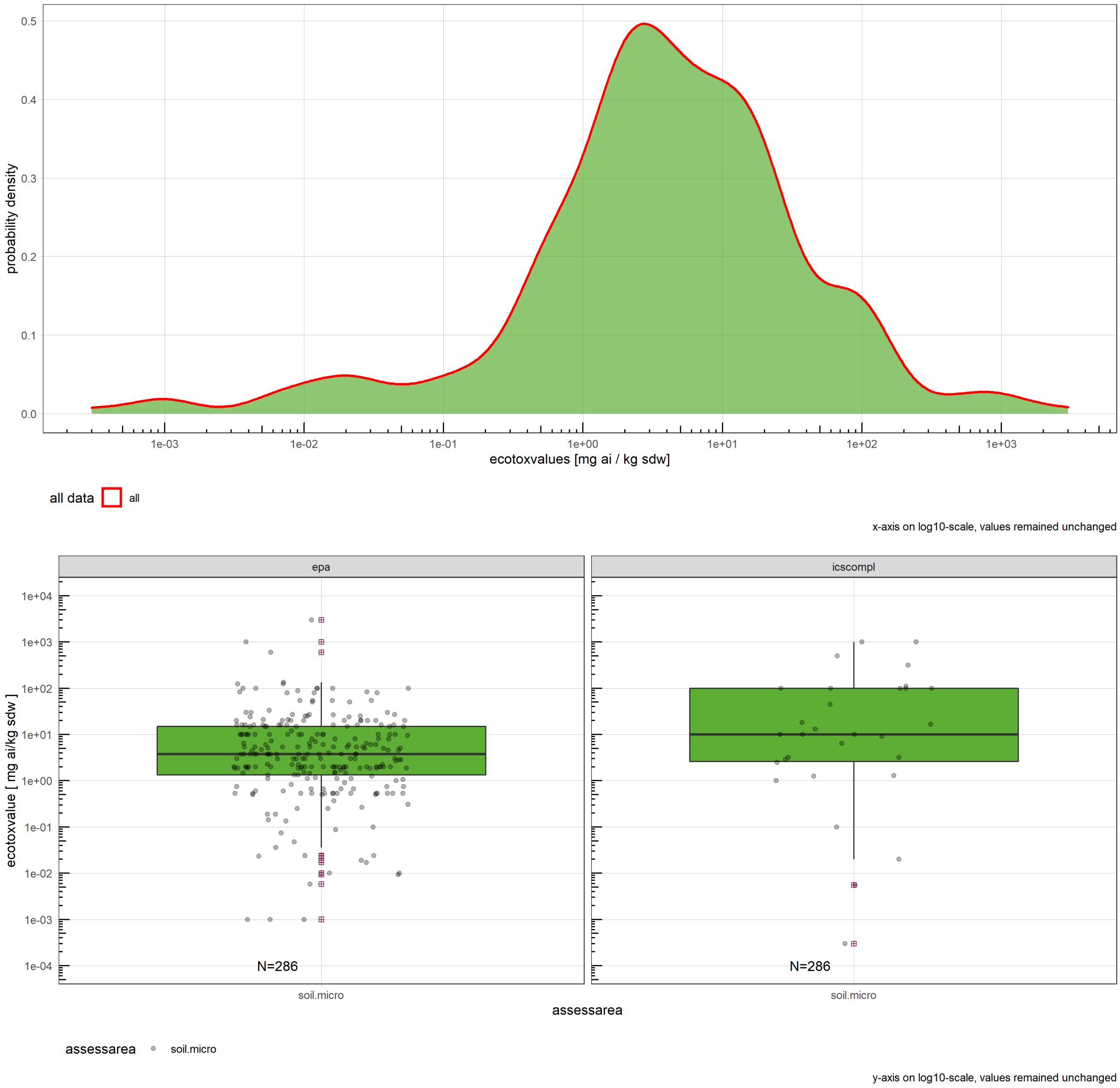
SUB: taxonomic subgroup of soil-macroorganisms; ORIG: the origin of the data as the data-providing repository (US EPA ECOTOX, UBA ETOX, IBA ICS, ECHA IUCLID); MEAN: arithmetic average of the ecotoxicological values of the criterion group; MEDIAN: median of the ecotoxicological values of the criterion group; GEO: the geometric mean of the ecotoxicological values of the criterion group; MIN: minimum of the ecotoxicological values of the criterion group; MAX: maximum of the ecotoxicological values of the criterion group; LQ10: lower 10th quantile of the ecotoxicological values of the criterion group; LU10: upper 10th quantile of the ecotoxicological values of the criterion group; N: sample size as the number of entries in the criterion group; all numbers except the sample size are given as mg ai / kg sdw

From the analyses of both acute and chronic toxicity distributions of soil macro-organisms, the ranges of ecotoxicological values were considered within a similar range, and most sub-group peaks (the highest probabilities for observing distinct toxicity values in a defined range around the modal value) were interpreted as overlapping. Except for nematodes and "other taxonomic subgroups", which were very rarely tested species groups, all distributions were considered sufficiently similar. All data for soil macro-organisms could be pooled, and no systematic difference between the taxonomic sub-groups was assumed.

5.3.5 Distribution of soil-micro chronic data – criterion group 5

The criterion group “soil-micro chronic” toxicity entries held very few data (Table 11, Figure 15).

Figure 15: Distribution of soil ecotoxicity values from chronic studies with soil-microorganisms (criterion group 5).



Source: own illustration, darwin statistics & gaiac

The density plot shows all values, and the boxplots show the distributions split into the data-supplying repositories.

Table 11: Statistical characteristics of the ecotoxicological values of criterion group 5.

ORIG	MEAN	MEDIAN	GEO	MIN	MAX	LQ10	LU10	N
epa	30.79	3.77	3.34	<0.01	3000.00	0.37	30.00	256
icscompl	119.36	10.00	7.86	<0.01	1000.00	0.06	408.00	30

ORIG: the origin of the data as the data-providing repository (US EPA ECOTOX, UBA ETOX, IBA ICS, ECHA IUCLID); MEAN: arithmetic average of the ecotoxicological values of the criterion group; MEDIAN: median of the ecotoxicological values of the criterion group; GEO: the geometric mean of the ecotoxicological values of the criterion group; MIN: minimum of the ecotoxicological values of the criterion group; MAX: maximum of the ecotoxicological values of the criterion group; LQ10: lower 10th quantile of the ecotoxicological values of the criterion group; LU10: upper 10th quantile of the ecotoxicological values of the criterion group; N: sample size as the number of entries in the criterion group; all numbers except the sample size are given as mg ai / kg sdw

5.3.6 Conclusion

In conclusion, the analysis of the distributions of ecotoxicity values for plants, soil macro-organisms, and soil microorganisms had shown sufficient similarity regarding their shapes and statistical characteristics within the five predefined criterion groups that pooling of data from different data sources and in the case of soil macroorganisms from taxonomic sub-groups was allowed in subsequent analyses. The five hypothesized criterion groups outlined above were used in the subsequent analyses to derive soil toxicity criteria.

6 Development of soil toxicity thresholds

At this point of the workflow and according to the conceptual approach (refer to section 1.6), the unified database of soil ecotoxicological information is called *prosoildat* and an operating handle (using R-scripts) for splitting the complete database into coherent criterion groups was available for further use.

Consistent, data-rich criterion groups for the derivation of soil toxicity criteria were built as described in the previous section for the three assessment areas "plant", "soil macro", and "soil micro". The most frequent ecotoxicological units were "mg ai / kg sdw" in combination with the statistical endpoints "NOEC/EC_{10}" or "EC_{50}". It was essential to split the analyses by either "acute" or "chronic" tests. The **five criterion groups** (refer to the preceding chapter) were used for explorative data analysis.

However, soil toxicity thresholds were developed using the same shared database for two different purposes.

- **CLP CLASSIFICATION:** Hazard classes serve the demands for classification and labeling according to the CLP Regulation (EC) 1272/2008. Substances are classified and labeled by different hazard classes (e.g., environment) and categories (e.g., toxic with long-lasting effects), indicating different levels of necessary precaution in, e.g., handling and transport.
- **PBT assessment:** The PBT criterion comprises criteria for persistency (P), bioaccumulation (B), and toxicity (T) which are defined under various chemicals regulations. Concerning active ingredients (PPP and biocides), PBT properties are a cut-off criterium, and these substances are excluded from approval/authorization and must not be marketed. REACH substances must be assessed for their PBT properties by registrants, handled appropriately, and identified as "substances of very high concern" (SVHC) by authorities.

Soil toxicity criteria related to chemical substances were derived by data-based tentative **toxicity thresholds** applying different methods and principles as outlined in the following.

6.1 Methods to determine toxicity thresholds

Acting on the assumption that the database *prosoildat* posed a representative sample of all conceivable ecotoxicity values for soil organisms, the left tail of the distribution pointed to chemicals that should be classified as very toxic. In contrast, other chemicals may still be considered toxic or at least harmful. Therefore, it was necessary to derive toxicity thresholds to have a scientific basis for setting criteria below which a substance could be classified as hazardous. As followed by ecotoxicological theory and as seen in most cases from real-life experiments, a response to a chemical exposure nearly never follows a pattern with a linear dose relationship, showing a clear threshold onsetting the effects. There is either a linear relationship with no threshold or even a decelerating relationship in most cases. The non-linear dose-response relationship means that the toxicity increases very fast at lower exposure concentrations and with a decreasing slope of the dose-response relationship at the higher doses. This scientific finding would have profound consequences in a hazard assessment of chemicals in that very low doses are proportionally more toxic (as shown for various substance classes such as benzenes, lead, radon) that nearly zero exposure would be the only protective measure for the environment and public health (Lanphear, 2017). Therefore, the precautionary principle is applied in current chemicals regulations (e.g., using assessment factors, specific measures for certain substances, or hazard-based assessments/regulations as for SVHC). Hazard assessments and legally binding threshold values aim to estimate low effect concentrations to protect the environment with various taxa and functions.

In contrast, criteria for classification & labeling purposes or PBT assessments conceptionally work in the opposite direction. Here, criteria are used to describe a substance's toxicity, "potency," or hazard and put it into a defined "class" with a defined upper threshold: the higher the toxicity thresholds were set, the more substances would fall into this class, reaching a higher protectiveness. The following basic assumptions (according to Carbonell et al., 1997) were applied here to enable a pragmatic and sufficiently protective approach.

- Each criterion group was defined as being inherently consistent and delivering comparable ecotoxicity values from representative taxa and the temporal aspect of a measurement endpoint according to the life cycle of the test organism.
- Each observation, i.e., each study, measurement endpoint, and test species, was equally important (as far as it belonged to the criterion group) and contributed to the finally used (possibly) aggregated ecotoxicological value.
- The statistical chosen approach should be as protective as possible in light of the evidence mentioned above and consequential restrictions for the environment, i.e., as often as possible and scientifically plausible, the toxicity value used for the ranking of a chemical should originate from the most sensitive species if several were available.
- It followed the principle of using a threshold that the fewer substances were classified as toxic, the lower the threshold.
- Three proposed thresholds needed to define
  - Which chemicals are very toxic?
  - Which chemicals are toxic?
  - Which chemicals are harmful?

The data analysis process, independently of the criterion group, ranked the substances by their toxicity in the first instance and then derived reliable thresholds for the hazard classes based on specific percentiles or other distribution-related measures.

The result of the data analysis process applying the thresholds to the ecotoxicity data was a **candidate list** of classified substances. Each classified chemical was assigned to one of six hazard classes, ordered by increasing concern for the environment: 1) harmful-acute, 2) toxic-acute, 3) very toxic-acute**,** and 4) *harmful-chronic*, 5) *toxic-chronic,* 6) *very toxic-chronic*.

Various authors described the methods of deriving reliable and transparent toxicity thresholds in the past (refer to the Background & introduction in chapter 1.3).

The authors used three approaches to describe the reference hazard class *"toxic":*

- The geomean approach,
- the quantile method and
- the Null Hypothesis Significance Testing (NHST) approach.

The primary rationale for using different statistical methods to define the toxicity thresholds was to empower the regulatory bodies to decide on the most protective approach, e.g., for the CLP regulation. The thresholds were set to percentiles of relatively high likelihood, e.g., the ten percent most toxic substances should be classified at least. It would also be possible to choose the fifth percentile for covering fewer chemicals or the twentieth percentile for covering more chemicals. The thresholds were complemented by freely modifiable factors or divisors applied to the original thresholds and shifted the base thresholds to determine the hazard classes "very toxic" and "harmful". This approach could be criticized as it may result in a high number of compounds being classified. It was not the purpose of this study to define a certain proportion of classified substances but to provide the scientific basis for toxicity ranges, derive soil toxicity thresholds, and propose suitable criteria to support decision-making instruments. The steps from data preparation to the description of the different approaches followed a strict order and intrinsic logic as outlined in the following.

6.1.1 Aggregation of datasets

The three approaches were applied to all five criterion groups *plants-acute*, *plants-chronic*, *macroorganisms-acute*, *macroorganisms-chronic,* and *microorganisms-chronic*. To model the distributions of ecotoxicological values and derive toxicity thresholds, it was necessary to use one unique value in cases where more than one ecotoxicological value was available for a particular substance. Therefore, an aggregation step preceded the data analysis. The most protective approach was selecting the lowest ecotoxicological value for every chemical substance. Quite many substances showed multiple entries (e.g., for the criterion group plants-chronic, the two substances atrazine and pendimethalin showed about 700 single observations each). The effect of aggregation was significant as the criterion groups were selected quite unspecific; the other way round, would the impact of data-rich substances be significant if data were not aggregated. The introduction of narrowly defined criterion groups made the effects of aggregation less pronounced.

6.1.2 Elimination of outliers

Outlying data points showing very high ecotoxicological values hampered the derivation of protective soil toxicity criteria in different manners. In cases where the distributions of ecotoxicological values contained many very high values ("right-tail outliers"), the resulting lower quantiles (e.g., the 10^{th} quantile) were broader compared to a dataset without outliers. The fact resulted in more substances being identified as candidate substances classified as "very toxic", "toxic", or "harmful", as well as fulfilling the T-criterion in a PBT assessment if applied to the method referred to as the "quantile approach" below. On the contrary, identifying candidate chemicals by applying the geometric mean (the approach and rationale of its use are described below) to the ecotoxicological values is less sensitive to outlying data points. Methods were available to remove outliers using the non-parametric Hampel test (Lehmann 2012, Dietrich & Schulze 2009). The algorithms developed here allowed for the choice of outlier removal. Comparisons for the five criterion groups had shown that many outliers would be detected if, e.g., the recommended Hampel criterion of five or the interquartile range method was applied. Up to one-third of the observations would have been skipped. In particular, plant data show (correctly, as comprehensible from the history and practice of plant testing) an extensive range of values. Variability across several orders of magnitude is also well-known for aquatic toxicity data.

However, it was decided that very high values were not considered outliers, as these less-toxic substances were believed to be an integral part of the whole "statistical population" of substances and thus cannot be discarded by a statistical technique alone. It followed from this fundamental decision that data were used without removing any outliers from the dataset. It could be discussed for future analyses if an alternative toxic threshold after outlier removal would deliver deviating results in terms of the total number of candidates for classification.

The approaches reflect different basic definitions of a hazardous substance. The **geomean approach** assumes that a substance has to be assigned as hazardous if it is *more toxic than an average substance of the statistical population of all substances*. The **quantile approach** assumes that a substance has to be assigned as hazardous if it is *more toxic than a distinct proportion of the total ranked population of substances*. The **NHST approach** assumes that a substance has to be assigned as hazardous if it is *significantly more toxic than the mean ecotoxicological value of the statistical population of all substances*.

6.1.3 The GEOMEAN approach

Renaud et al. (2004), also referred to as the "EU concept", used acute laboratory ecotoxicological values from earthworm studies to calculate the geometric mean. Generally, the *geometric mean* must be used for log-normally distributed data (i.e., it followed a normal distribution after converting the measurement values by the natural logarithm). They divided the resulting geomeans by 100 to derive a cut-off value for "very toxic" substances. No rationale, no reference, or resulting protection levels were given. However, the method was recommended for the classification of potentially hazardous substances. The same approach was followed because of the right-tailed distribution of ecotoxicological values observed for most data groups. The method was also recommended by (the recently updated) OECD GD 23 to calculate mean exposure concentrations for the testing of difficult substances and mixtures that showed log-normally distributed ecotoxicological values (OECD 2000) and was here applied accordingly. To assure the reliability of the values, all censored data, i.e., containing ">" operators (i.e., the highest concentration tested did not show any significant effects), was deleted before the analysis. The problem of using or not using censored values was also stressed in chapter 5.1.6. An operator was available for about 16 % of the data entries, 10 % of which were classified as ">", amongst nine other operator categories ("=", ", ">", ">=", "~", "<=", "ca.", "<<", "ca.", "=>").

In the present approach, the geometric mean marked the upper limit of the "toxic" hazard class. The threshold values for "very toxic" and "harmful" hazard classes were deduced by applying additional multipliers or divisors to the base threshold. It was decided to use a multiplier of ten to derive the "harmful" category and a divisor from deriving the "very toxic" class. The current approach deviated from the historical approach of Renaud et al. (2004). Theoretically, it led first to more very toxic substances, second to more toxic substances, and third to the additional assignment of harmful substances (an assignment for harmful substances was not undertaken by Renaud et al., 2004). For visualization, a "number ray" setting the borders of the hazard classes could be imagined. It was more likely that a substance would be classified at least as harmful by the current approach. In a highly complex workflow like the one at hand, there are manifold (arbitrary or challenging to explain decisions) to be taken that significantly influence the analyses' outcome.

The geometric mean is defined by n^{th} root of the product of n observations (while the arithmetic mean uses the mean sum). It shows a typical value of a set of numbers, the central tendency, and is calculated after Equation 1 and concretely using the alternative way of Equation 2.

Equation 1: Computation of the geometric mean by the root method

$$\bar{x}_{geo} = \sqrt[n]{x_1, x_2 \dots x_n} = \left[\prod_{i=1}^{n} x_i\right]^{1/n}$$

Equation 2: Computation of the geometric mean by the logarithm method

$$\bar{x}_{geo} = exp\left[\frac{1}{n}\sum_{i=1}^{n} \log(x_i)\right] = e^{\bar{y}}, \text{with } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_j \times y_i = \log x_i,\ i = 1,2,\dots$$

The method was implemented in an adapted function. However, the results and methodology were equivalent to Millard (2013).

The use of the geomean posed an approach that built the ratio of an "average" toxicity characterized by the central tendency of the distribution, the geomean, and a fixed factor, here ten, to mark the proportion of which substances were deemed much more or much less toxic than an average chemical. The factor of ten was chosen to deviate from the EU concept of Renaud et al. with a factor of 100 to avoid constructing a very broad class of harmful substances.

6.1.4 The QUANTILE approach

The quantile approach was proposed by Hartmann et al. (2014), following Carbonell et al. (1997) and Torstensson et al. (1999). They used the lower 10^{th} quantile as the threshold to separate the most toxic substances from the overall distribution of ecotoxicological values for all substances. Carbonell et al. used data from about 1000 substances to derive hazard classification criteria for terrestrial organisms. They differentiated taxonomic groups (soil organisms, microorganisms, bees, plants) and added criteria such as biodegradation, mobility of a substance in soil, and other exposure indicators (duration) in a combined matrix of hazard classes.

Here, the lower 10^{th} quantile of the criterion group substances from a kernel density probability estimates independent of the underlying distribution (e.g., no normal distribution was a prerequisite) was assigned to be at least "toxic". The hazard classes "very toxic" and "harmful" were deduced from applying multipliers (includes harmful substances) or divisors of two (separates very toxic substances) to the threshold values derived by the quantile method. The application of the factors resulted in thresholds for very toxic substances at approximately the 5^{th} lower quantile and the harmful substances at the lower 20^{th} quantile, respectively.

Quantiles of the underlying distributions of ecotoxicological values were calculated using the base functions in R language, following the recommendations of Hindman & Fan (1996), defining the type = 8 that gains a median-unbiased estimator independent of the distribution (out of nine methods available in the base-R quantile function).

6.1.5 The NHST-approach

An alternative deriving hazard thresholds for ecotoxicological values was to apply a Null Hypothesis Significant Test (NHST) for differences between the means of 1) the distribution of a subset of the data for a distinct substance and 2) the remaining data for all the other substances minus the focused substance. The main difference between the two former approaches (quantile and geomean) was the **inclusion of all available data**, incorporating the variability of all measurements for a specific substance and not filtering for the most sensitive observations per substance. The risk was that the results of the three approaches were not well comparable. The number of classified substances depended on the selected approach and primarily on the additional factors that should account for the uncertainty of the definition of thresholds and could be aligned so that all approaches deliver a similar amount of chemical classified. The NHST-approach was applied to each criterion group (out of five), applying a one-sided smaller Welch t-test that allows for unequal sample sizes and unequal variances from the base R-package "stats" (R Core Team 2020). A stepwise approach was developed and applied to the data points of each test substance within the five criterion groups.

- The subset of criterion group data was split into *testdata* (only entries from the substance focused) and *compdata* (all substances without the substance focused).
- A one-sided smaller Welch t-test (Welch 1947) on differences between the means of log-*testdata* and log-*compdata* was applied. The pairwise t-Test after Welch adapts the Student´s t-test to situations where the two variances of the statistical populations are unequal. The prerequisite of normally distributed values is still given. Pre-tests for normality (Shapiro-Wilk test, Shapiro & Wilk 1965) were only necessary if sample sizes were ≤ 5000. Otherwise, normality could be assumed from the Central Limit Theorem of statistics. Homoscedasticity checks were conducted using the Levene test (Levene 1961). Pre-checks were not used as exclusion criteria. Proposing the NHST-approach was considered an unorthodox alternative that would have to be discussed intensively before being used by default for the derivation of thresholds for CLP. Non-parametric test procedures to detect the significance of differences in a single chemical's toxicity compared to the average of all available data (within one of the five criterion groups) could be easily implemented in an approach that would be more elaborated in the future. It was decided to use the adapted t-test procedure regardless of the results of the pre-testing for normality because the sample sizes of the test data were sufficiently high to assume normality regardless of the Shapiro-Wilk-test result, and the data pairs should be only excluded if both pre-tests ("more than fifty percent of the tests") would fail. A sufficient number of significant results should inform the comparison of the three approaches rather than providing a readily elaborated methodology that considers all eventualities of a data structure. Table 12 shows that a considerable number of pair-wise comparisons would be excluded if the normality criterion for the large test dataset had been applied, namely the combination EXCLU, i.e., the number of substances to be excluded from the derivation of the NHST-threshold because as well the normality check for *testdata* and *compdata* failed for the same substance. It could be discussed if a formal statistical test on the relatively large sample sizes of the *compdata* (refer to the total number of entries for each criterion group from Figure 11, Figure 12, Figure 13, Figure 14, and Figure 15).
- The p-values were listed for the data subsets *compdata* for which calculations could be conducted given sufficient sample sizes of the focused substance.
- The list of p-values was ordered ascending.
- At the first non-significant p-value ($p \geq 0.05$), the list was cut, and the geomean of the *ecotoxvalue* of all substances below the significance threshold of $p < 0.05$ was set as the NHST-threshold for the hazard class "toxic". The approach aimed at reflecting the range of ecotoxicological values that appeared to be significantly lower than an average substance, and in this way, the approach resembled the GEOMEAN approach.

Table 12: Overview of pre-testing results for normal distribution and variance homogeneity.

GROUP	SIG [TOT]	NORM FAIL TEST	NORM FAIL COMP	EXCLU	HOMSCED FAIL
Plants acute	76 [339]	53	76	53	7
Plants chronic	238 [677]	0	18	0	234
Soil-macro acute	31 [508]	16	31	16	9
Soil-macro chronic	55 [546]	26	55	26	33
Soil-micro chronic	7 [99]	5	7	5	6

GROUP: five criterion groups holding comparable datasets; SIG [TOT]: number of significantly different chemicals [out of the total number of substances within the criterion group], (p-value < 0.05); NORM FAIL TEST: the number of significantly different substances that did not pass the normality test for the *testdata* (single chemicals) (p-value < 0.05); NORM FAIL COMP: the number of significantly different substance entries for which the normality check of complete dataset *compdata* without the test substance focused failed (p-value < 0.05); EXCLU: number of substances to be excluded from the derivation of the NHST-threshold because as well the normality check for *testdata* and *compdata* failed for the same substance; HOMSCED FAIL: number of significantly different substances for which not variance homogeneity could b assumed.

The hazard class "harmful" was derived by multiplying the NHST-threshold by two and the hazard class "very toxic" by dividing the NHST-threshold by two. The factor was chosen based on the consideration that a reliable amount (and not the vast majority) of substances should be classified.

6.1.6 The SSD approach determined a probabilistic value for each substance

The Species Sensitivity Distribution (SSD) is a theoretical and statistical approach to describe the relative sensitivity of tested species and to determine a safe concentration of a single chemical substance for most of the species naturally occurring in an environmental compartment or habitat, represented by surrogate species that help to build the SSD (e.g., Aldenberg & Jaworska 2000). It takes the form of a dose-response curve with an indicator of toxicity and sensitivity of a species towards the chemical, e.g., EC_{50}, on the log-scaled x-axis versus the relative fraction of the test population of single species after ordering them by their relative sensitivity. The uncertainty of SSD for risk assessment that aims to define a safe (soil) concentration or application rate is usually described as the upper bound of the 95 % confidence interval of the fraction of species potentially affected (PAF). On the other hand, the uncertainty of an SSD for the extrapolation to (theoretically all) untested situations is described as the 95 % hazardous concentration at which five percent of the available species were affected, the HC_5. The uncertainties depend on the number of species available for computing the SSD and the shape and strength of the dose-response dependent distribution of ecotoxicity values. At present, the uncertainty originated from the fact that only a few standard species were tested, and not necessarily the most sensitive species were of paramount interest.

SSDs were modeled for all substances with sufficient data (see below) within the five criterion groups. They are of two-fold use:

1. Complementary to the minimum ecotoxicological values used for the quantile and geomean approach, a modeled measure of toxicity for each of the substances within a criterion group could be included in the evaluation that models the distribution of sensitivities of all species tested. This approach included more information than the former approaches (like the NHST approach): the concentration of a hazardous substance for 5 % of the species – $\mathbf{HC_5}$.
2. SSDs showed how often the suspected most sensitive species, which should be the standard test species as a protective proxy of the toxicity of a substance, were indeed the most sensitive taxa. In many cases, it could be demonstrated that non-standard species were more

sensitive than the standard species. It was concluded that a factor of extrapolation from few to many species was principally barely constant and that any species tested was helping in reducing the uncertainties remaining in risk- or hazard-based environmental assessment.

SSDs were fitted applying a log-normal distribution to the *ecotoxvalue* vs. fraction affected data (Wheeler et al., 2002, fit by the *fitdistrplus*-package in R, Delignette-Muller & Dutang, 2015). The goodness-of-fit of the data with log-normal distribution was assessed using the Anderson-Darling test (e.g., Stephens, 1986, $p < 0.05$). From the bootstrapping algorithm that computed the confidence intervals (parametric bootstrapping using the fitted distribution to sample, 1000 iterations), the convergence of each step was noted. An SSD was only interpreted and reported if the two quality criteria, goodness-of-fit, and convergence, were passed. The aggregation of multiple entries for each combination of a species and a substance was done using the geomean (not the overall lowest value was used for the quantile and the geomean approach). Data from at least **five** species (Table 16) had to be available within each criterion group; detailed results are provided to the contractor UBA in the supplementary material. The criterion was chosen to be less strict than usually recommended in the literature (Wheeler et al. recommended 10-15 values) because the authors considered the rank of a species more important than the reliability of the HC_5.

6.2 Deduction of toxicity thresholds

Separate toxicity thresholds based on the geomean-, the quantile- and the NHST-approaches were calculated for the five criterion groups, as defined in section 6.1.

The geomean method proved consistently to be more conservative or protective and would lead to a high number of classified substances. In that sense, it produced **higher thresholds** of the ecotoxicity values (Table 13). Consequently, the list of potential candidates classified as toxic was longer for the geomean approach (the relevance and regulatory consequences will be discussed in chapter 7). The number of classified substances in the category "toxic" was for the *geomean* approach between 3-times (plants acute and soil macroorganisms acute) and 7-times higher (soil macroorganisms chronic) than for the *quantile*- and *NHST*-approaches. The distance between measures of central tendencies like arithmetic or geometric means and the probability of seeing a range of values as for the quantile approach depends mainly on the skewness of the distributions. The distribution of ecotoxicological values for the effects of (toxic) chemicals on soil organisms was mainly right skewed, meaning that there were many small and few high values. The thresholds and resulting base categories mainly reflected different levels of protectiveness between the three approaches. The NHST and the quantile approach gave similar numbers of classified substances. The number of substances classified for the quantile and the NHST approach was similar, except for group five (i.e., soil microorganisms chronic). Please be aware that the sample size in Table 13 refers to the aggregated data, i.e., only the lowest ecotoxicological value for each substance was taken.

Table 13: Toxicity threshold values (units in mg ai / kg sdw) and the number of classified substances within the five criterion groups. Thresholds are given for the base toxicity class "toxic".

criterion group	Group 1 plant acute	Group 2 plant chronic	Group 3 soil-macro acute	Group 4 soil-macro chronic	Group 5 soil-micro chronic
geomean threshold	1.615	0.807	55.267	25.889	1.973
quantile threshold	0.001	0.001	0.376	0.500	0.111
nhst threshold	0.177	0.143	0.787	0.770	0.667
total no of substances in group	339	677	506	546	99
no substances classified geomean	34	118	63	157	40
no substances classified quantile	10	19	20	23	2
no substances classified nhst	11	23	15	24	20

The threshold values for the central hazard class "toxic" and the corresponding number of substances affected by the specific threshold were shown.

6.3 Candidate lists of classified chemicals

After fixing the threshold values for the class “toxic” and applying the additional approach-specific factors to the thresholds that define the “very toxic” and ”harmful” classes, the specifications were practically used within each of the five criterion groups. The result was a “*candidate list*” of all potentially classified substances for any of the three approaches in any of the three toxicity classes. The candidate list contained all substances that fulfilled at least one classification criterion. It could be without further ado that all of the three approaches classified the same substance, and it is further possible that from each of the three approaches, a different toxicity class was assigned to the same substance. A candidate list held information on the identity of the classified substance (*prosoildat* variables *aicas*, *aicommon*), the regulatory context as the variable *regarea* set, the number of data entries of the respective substance in the respective criterion group, and its aggregated (to the minimum of all available values) *ecotoxvalue*. The thresholds were applied, and the classes were assigned and marked within the candidate list as the “rank_geomean”, “rank quantile,” and “rank_sigtest” which classifier applied to which substance in the list that was sorted by the lowest *ecotoxvalue*. “verytoxic11” from column “rank_geomean” was the 11th-toxiest substance according to the geomean approach and similar for all substances classified and approaches used. The lists did not show substances that were not classified at all.

Additionally, the most sensitive species and external classifiers from ECHA lists on the status of PBT, SVHC, POP, and CL were given. An example is shown for the actual candidate list for criterion group 4, “soil macroorganisms chronic”, with the ecotoxicological unit "mg ai / kg sdw" and the statistical endpoint NOEC. In-deep analyses of comparisons between the PROSOIL and existing classifications are shown and discussed in chapter 7. The candidate list derived from soil toxicity thresholds from chronic data on soil macroorganisms (Table 14) comprised pesticides predominantly; 21 out of 30 most toxic substances were regulated under the legislation for plant protection products (although some of these substances no longer approved as actives ingredients in PPP). The dominance of pesticides amongst the candidates was not surprising because this class of substances was designed to be toxic enough to control at least the target pest species in the field, and the standard species in the laboratory should be at least susceptible as suitable indicators of hazard.

The three data analysis approaches to derive soil toxicity thresholds deliver concordant candidates within the 30 most toxic substances, differences between the various methods were seen beyond the most toxic end of the ranking. Most of the most sensitive species were recruited from the standard test species. The fact that these standard species were by no means the most sensitive species in diversified species sensitivity distributions, as demonstrated in chapter 3.6.3, could not be apparent in a candidate list. The standard species were tested much more frequently and with almost every substance; on the other hand, non-standard species were studied rarely. The available data points to derive the most sensitive data point was found by first aggregating several ecotoxicological values by the minimum value for each species and second by taking the minimum value out of the data for all species. The range of available values amongst the 30 most toxic substances was between 1 and 100. Additional candidate lists for the remaining four criterion groups (except group four) can be found in the supplementary material (A.2 annex A.2.). Detailed data was made available to UBA.

Table 14: Candidate list of criterion group 4 soil macro chronic, 30 most toxic candidates

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
10108-64-2	100	≤ 0.001	FC	cadmium chloride	verytoxic01	verytoxic01	verytoxic01	REACH		SVHC		h400 & h410	h400 & h410
2921-88-2	36	≤ 0.01	AC	chlorpyrifos	verytoxic02	verytoxic02	verytoxic02	PESTI	listed			h400 & h410	h400 & h410
79622-59-6	1	≤ 0.01	EF	fluazinam	verytoxic03	verytoxic03	verytoxic03	PESTI				h400 & h410	h400 & h410
10605-21-7	84	≤ 0.01	EA	carbendazim	verytoxic04	verytoxic04	verytoxic04	BIOC&PESTI				h400 & h410	h400 & h410
210880-92-5	11	≤ 0.01	EF	clothianidin	verytoxic05	verytoxic05	verytoxic05	CROSS&PESTI				h400 & h410	h400 & h410
60-51-5	30	≤ 0.01	FCa	dimethoate	verytoxic06	verytoxic06	verytoxic06	PESTI				h400	h400
208465-21-8	2	≤ 0.1	EF	mesosulfuron-methyl	verytoxic07	verytoxic07	verytoxic07	PESTI				h400 & h410	h400 & h410
13194-48-4	4	≤ 0.1	FC	o-ethyl s,s-dipropyl phosphorodithioate	verytoxic08	verytoxic08	verytoxic08	LITRA				h400 & h410	h400 & h410
1668-54-8	14	≤ 0.1	FC	4-methoxy-6-methyl-1,3,5-triazin-2-amine	verytoxic09	verytoxic09	verytoxic09	CROSS					
41198-08-7	1	≤ 0.1	FC	profenofos	verytoxic10	verytoxic10	verytoxic10	REACH				h400 & h410	h400 & h410
135285-90-4	18	≤ 0.1	EF	hexanitrohexaazaisowurtzitane	verytoxic11	verytoxic11	verytoxic11	LITRA					
135410-20-7	3	≤ 0.1	EF	acetamiprid	verytoxic12	verytoxic12	verytoxic12	PESTI&BIOC				h400	h400 & h412
70288-86-7	33	≤ 0.1	PM	ivermectin	verytoxic13	verytoxic13	verytoxic13	VETME					h400 & h410

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
81-07-2	4	≤ 0.1	EF	1,1-diox-1,2-benzisothiazol-3-one	verytoxic14	verytoxic14	verytoxic14	CROSS					
83121-18-0	1	≤ 0.1	FC	teflubenzuron	verytoxic15	verytoxic15	verytoxic15	PESTI					h400 & h410
120068-37-3	3	≤ 0.1	FC	fipronil	verytoxic16	verytoxic16	verytoxic16	PESTI&REACH &BIOC				h400 & h410	h400 & h410
71751-41-2	26	≤ 0.1	EF	abamectin	verytoxic17	verytoxic17	verytoxic17	PESTI&BIOC				h400 & h410	h400 & h410
138261-41-3	15	≤ 0.1	FC	imidacloprid	verytoxic18	verytoxic18	verytoxic18	PESTI&REACH &VETME&BIOC				h400 & h410	h400 & h410
1912-24-9	8	≤ 0.1	EA	atrazine	verytoxic19	verytoxic19	verytoxic19	PESTI				h400 & h410	h400 & h410
5755-27-1	5	≤ 0.1	EF	1,3-Dinitro-5-nitroso-1,3,5-triazinane	verytoxic20	verytoxic20	verytoxic20	LITRA					
61213-25-0	1	≤ 0.1	EF	flurochloridone	verytoxic21	verytoxic21	verytoxic21	PESTI				h400 & h410	h400 & h410
87-86-5	19	≤ 0.1	EA	pentachlorphenol	verytoxic22	verytoxic22	verytoxic22	PESTI				h400 & h410	h400 & h410 & h413
91465-08-6	5	≤ 0.1	PP	lambda-cyhalothrin	verytoxic23	verytoxic23	verytoxic23	PESTI&REACH &BIOC				h400 & h410	h400 & h410
114-26-1	1	≤ 1	EF	propoxur	verytoxic24	verytoxic24	verytoxic24	PESTI				h400 & h410	h400 & h410
7447-39-4	93	≤ 1	FF	copper-II-chloride	verytoxic25	verytoxic25	verytoxic25	REACH					h400 & h410 & h411
74223-64-6	1	≤ 1	EF	metsulfuron-methyl	verytoxic26	verytoxic26	verytoxic26	PESTI				h400 & h410	h400 & h410

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
112281-77-3	3	≤ 1	EF	tetraconazole	verytoxic27	verytoxic27	verytoxic27	PESTI				h400	h411
119446-68-3	7	≤ 1	EF	difenoconazole	verytoxic28	verytoxic28	verytoxic28	PESTI					h400 & h410
136426-54-5	1	≤ 1	EF	fluquinconazole	verytoxic29	verytoxic29	verytoxic29	PESTI				h400 & h410	h400 & h410
17804-35-2	23	≤ 1	EF	methyl 1-(butylcarbamoyl)benzimidazol-2-ylcarbamate	verytoxic30	verytoxic30	verytoxic30	PESTI				h400 & h410	h400 & h410

CAS = CAS number; N = n data; VAL = ecotox value rounded [mg ai/kg sdw]; SENS = most sensitive species; NAME = common name; RGEO = rank geomean approach; RQUA = rank quantile approach; RSIG=rank sigtest approach; REG = registration area; PBT = PBT status; SVHC = SVHC status; POP = POP status; CLHAR = ECHA CL status harmonized; CLSEL = ECHA CL status self-classified; species names abbreviated: FC = Folsomia candida; AC = Aporrectodea caliginosa; E = Eisenia fetida; EA = Enchytraeus albidus; PM = Proisotoma minuta; PP = Porcellionides pruinosus; FF = Folsomia fimetaria

6.4 Comparability of the approaches

After determining the toxicity thresholds and looking at the potential consequences of applying the derived soil toxicity criteria to the underlying data ("the candidate lists"), an obvious question was: "How comparable are the toxicity thresholds, e.g., if new data would be available or if stricter quality criteria would lead to fewer compounds classified?" Hundred random sub-samples of varying sizes (between a minimum value of 50 observations and the original size of the dataset) were drawn from the original data set to answer this question. The toxicity thresholds for the three approaches and within each criterion group were calculated for each new dataset. The uncertainty of the derivation of the toxicity thresholds was then given as the upper and lower 95 %-confidence intervals of the mean threshold, which signifies that a mean threshold would be in 95 of 100 cases of using randomly smaller datasets in the range of the confidence interval. Please be aware that the mean sample size in Table 15 referred to the aggregated data, i.e., only the lowest ecotoxicological value for each substance was taken. The HC_5 could not be used for this analysis because each of the chemicals within a criterion group could not fulfill the quality and goodness of fit criteria for computing an SSD (6.1.6).

Table 15: Stability of the methods of deriving toxicity thresholds (class "toxic" threshold).

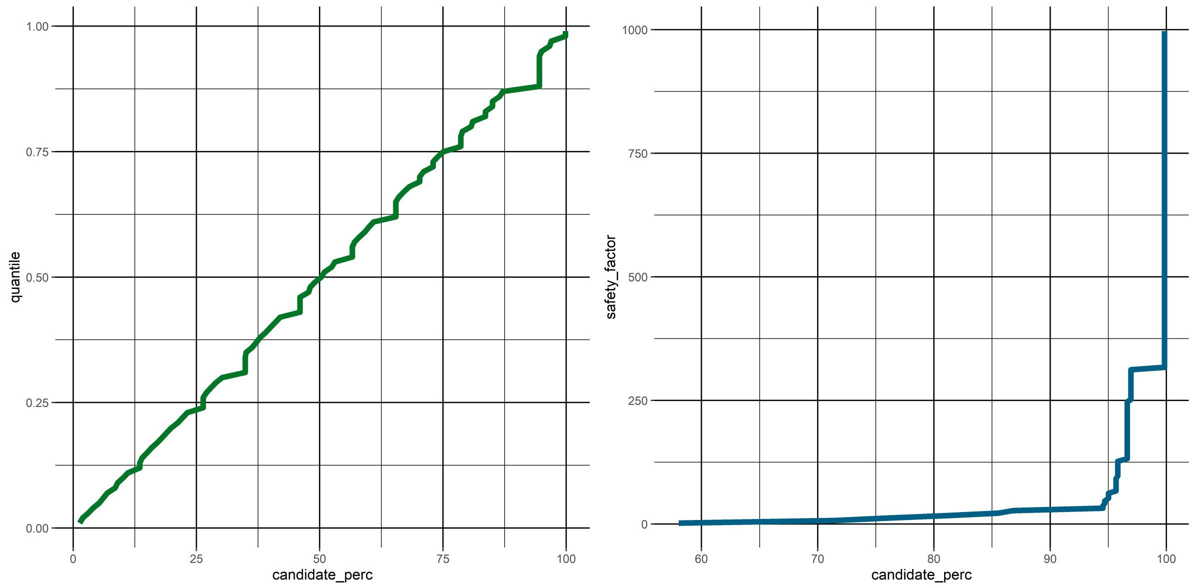
Group	Average N	GEO Mean	GEO CI_{low}	GEO CI_{high}	QUA Mean	QUA CI_{high}	QUA CI_{high}	NHST Mean	NHST CI_{high}	NHST CI_{high}
1	180	1.722	1.614	1.830	0.002	0.001	0.002	0.341	0.288	0.393
2	343	0.870	0.816	0.925	0.001	0.001	0.002	0.268	0.251	0.284
3	292	54.763	52.850	56.677	0.434	0.394	0.473	2.117	1.675	2.559
4	301	25.468	24.541	26.396	0.537	0.506	0.567	3.608	3.249	3.967
5	77	2.057	1.986	2.129	0.154	0.134	0.173	0.774	0.656	0.891

Average N: Mean sample size of the 100 random samples used to compute thresholds. Means and upper/lower confidence intervals (CI_{high}, CI_{low}) were computed based on 100 independent random samples of the original datasets. Group: One of the five criterion groups "1 = plant-acute", "2 = plant-chronic", "3 = soil-macro acute", "4 = soil-macro chronic", "5 = soil-micro chronic". GEO = Geomean; QUA = quantile

The randomization runs (see the "complete threshold table" from the supplementary material) showed that the threshold values would change to a greater extent if the database would change clearly and to a greater extent. The geometric means used as threshold values (i.e., showing the limit to "toxic" substances), derived from 100 independent random samples, were often similar but not identical to those derived from the original data. The confidence intervals around the means were narrow, indicating that the methods do not depend strongly on the number of observations if the database was large enough. The stability exercise accounted for smaller datasets only because a random subset of the existing data was used. It was impossible to "upscale" the available data to larger sample sizes.

The number of candidates was differently related to the approach chosen to deduce thresholds. Suppose the quantile was plotted against the relative share of substances classified as all substances within a criterion group (as in Figure 15 for the criterion group 4). In that case, the quantile approach has a constant slope, i.e., if you choose a particular quantile of distribution, you get a certain percentage of a list of substances. On the contrary, the relative share of the total list for the geomean approach starts at more than 50 % of the substances already classified as "toxic" without any additional assessment factor. If you divide the geomean by additional factors, e.g., ten, then the geomean approach would classify a proportion of about 80 % of all substances.

Figure 16: Differences in the mechanisms that led from two methodological approaches, the "quantile" (left) and the "geomean" (right), to the assignment of a certain percentage of the total list of chemicals within a criterion group (here, group 4).



Source: own illustration, darwin statistics & gaiac

6.5 Most sensitive or frequently studied?

The chemical legislations require the testing of important groups of soil organisms, providing the biodiversity and functions of soils, e.g., predatory and decomposer species to different extents (EFSA 2017 from the plant protection products´ perspective), e.g., the soil microflora (fungi and bacteria) was only addressed using a functional test. The measurement endpoints were nitrogen- and organic carbon transformation rates, and the trigger for "no observed effects" were defined as "no larger deviation of the (single) treatment group from the control group than 25 % after 100 days of exposure" (methods after OECD 2000A, B).

Few methods according to ISO or OECD standards comprising a limited number of different soil organism species have been established in recent years (Alves & Cardoso 2016). In environmental risk assessment for soil organisms, the taxonomic group of earthworms was the most frequently studied. Mainly the compost worm species *Eisenia fetida* or *Eisenia andrei* (refer to section 6.1.6, after OECD guideline 222, OECD 2004), including other species from the group of oligochaetes like the enchytraeids *Enchytraeus albidus* and *Enchytraeus crypticus* (according to guideline 220, OECD 2016) were tested. Other standardized tests were available for mollusks (*Helix aspersa*), mites (gamasid *Hypoaspis aculeifer*, oribatids *Platynothrus peltifer,* and *Oppia nitens*), isopods (*Porcellio scaber* and *Porcellionides pruinosus*), and collembolans (*Folsomia candida* and *Folsomia fimetaria*).

For the environmental risk assessment in different regulations and hazard classification schemes, the standard test species represented a trade-off between the most feasible species for testing and a sensitive surrogate species for a taxonomic or functional group.

As the *prosoildat* database integrated standard and non-standard study results from various sources, it became possible comparing the sensitivity ranking among the most studied species. From this analysis, conclusions could be drawn if the "standard test battery" is likely sufficiently protective to cover most hazards from the exposure to chemicals. Furthermore, the HC_5-values could be alternatively used for this analysis (section 6.1.6).

Examples of the results of the computation of the SSDs that fulfilled the quality criteria were selected from **criterion group 4,** which comprised chronic studies with soil macroorganisms. The criterion group was considered relevant in potentially containing the most relevant test protocols for soil chronic hazard assessments. In total, 28 SSDs were computed for substances comprising sufficient data points (at least five) and for which the modeling was successful in terms of model validity (see Table 15). All SSDs shown here converged from iteratively determining bootstrap confidence intervals based on the underlying distribution model and were valid according to the goodness-of-fit measure (method described in chapter 6.1.6) out of 546 substances within the criterion group 4.

Most taxonomic data were available for the substance *carbendazim* (see Table 16). The fungicidal substance carbendazim (which was contained in the plant protection formulation "Derosal") was the positive control substance for many laboratory and field studies with earthworms due to its high toxicity towards earthworms, similar to benomyl. The carbendazim dataset contained twenty species, including higher taxonomic entities, like Annelida or Collembola, used to compute the SSD analysis. The inclusion of higher taxonomic entries was questionable because of ambiguities that could occur. However, SSDs were not meant to be used directly for regulatory purposes but to analyze the relative sensitivity of the standard species, so the higher taxonomic units were to be considered when looking for striking patterns of the SSDs.

As shown in figure 16, the most sensitive species was the enchytraeid *Enchytraeus albidus*, with a geomean of all NOEC values (0.39 mg ai/kg sdw), which is well below the reported HC_5 of 0.58 mg ai / kg sdw, but not below the lower credible interval of 0.17 mg ai / kg sdw (Figure 17).

On the other hand, not always the "usual suspects (i.e., the standard test species) ranked at the lower tail of the sensitivity distribution. For example, for the many copper compounds that were systematically tested in the past, the most sensitive species was the nematode *Plectus acuminatus*.

Two use cases of SSD results could be envisioned:

- Firstly, the last time Frampton et al. (2006) compiled soil data systematically to compute SSDs. Since *prosoildat* offered novel data compilation, SSD-specific indicators of protective thresholds, e.g., the HC_5 for each chemical, could be used for a generic classification of chemicals and a toxicity ranking. This generic classification could derive toxicity thresholds for classification and labeling by introducing additional safety factors and deriving cut-off values for PBT assessments.
- Secondly, it could be determined which species or groups were frequently among the most sensitive species. From this could be generally concluded that stipulating the development of standard protocols for new test species would be recommendable, or the introduction of additional safety factors in cases where only relatively insensitive species were tested would be necessary.

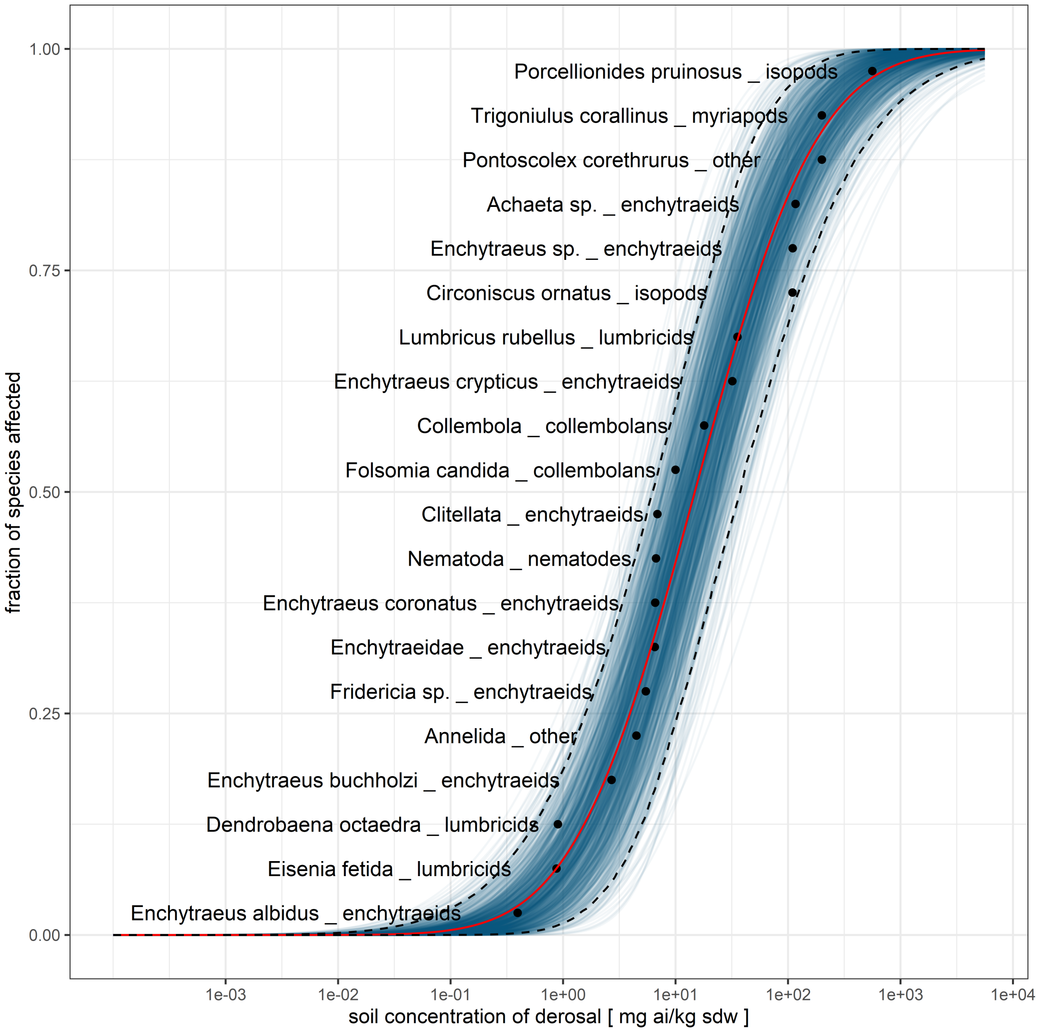
Table 16: Species sensitivity distributions of 20 most species-rich chemical substances (historically, copper compounds were studied extensively, refer to Frampton et al. 2006) within the criterion group 4 soil macro chronic.

CAS: unique identifier CAS number; NAME: common name of the chemical compound; HC5: Hazard concentration for 5% of the population affected; UPPER: upper 95 % quantile of HC5; LOWER: lower 95 % quantile of HC5; MOSTSENS: most sensitive species contributing to the respective SSD; NOSPEC: number of species included in the analysis.

CAS	NAME	HC5	UPPER	LOWER	MOSTSENS	NOSPEC
10605-21-7	carbendazim	0.58	2.18	0.17	*Enchytraeus albidus*	20
7447-39-4	copper dichloride	37.69	70.75	22.11	*Plectus acuminatus*	16
7758-98-7	copper sulphate	22.42	68.85	8.57	*Plectus acuminatus*	16
7440-50-8	copper	26.34	66.2	12.89	*Aporrectodea caliginosa*	14
70288-86-7	ivermectin	0.04	0.59	0.01	*Proisotoma minuta*	12
3251-23-8	copper dinitrate	22.02	58.98	10.39	*Aporrectodea caliginosa*	10
12069-69-1	basic copper carbonate	21.67	62.84	10.23	*Aporrectodea caliginosa*	9
12158-74-6	copper (ii)-hydroxyphosphate	21.67	64.7	9.11	*Aporrectodea caliginosa*	9
1317-38-0	copper oxide	21.67	67.34	9.31	*Aporrectodea caliginosa*	9
1317-39-1	dicopper oxide	21.67	65.95	10.21	*Aporrectodea caliginosa*	9
142-71-2	copper acetate	21.67	65.28	9.15	*Aporrectodea caliginosa*	9
17804-35-2	benomyl	0.41	3.36	0.08	*Eisenia fetida*	9
20427-59-2	copper hydroxide	21.67	64.04	8.97	*Aporrectodea caliginosa*	9
22205-45-4	dicopper sulphide	21.67	63.76	8.62	*Aporrectodea caliginosa*	9
60-51-5	dimethoate	0.54	1.92	0.21	*Enchytraeus albidus*	9
7681-65-4	copper iodide	21.67	62.48	9.5	*Aporrectodea caliginosa*	9
7758-89-6	copper chloride	18.22	72.12	6.28	*Aporrectodea caliginosa*	9
2921-88-2	chlorpyrifos	0.24	1.91	0.05	*Folsomia candida*	8
10108-64-2	cadmium chloride	6.19	20.1	2.63	*Folsomia candida*	7
71751-41-2	abamectin	0.4	2.08	0.11	*Folsomia fimetaria*	6

All computable SSDs have been made available to UBA.

Figure 17: Species sensitivity distribution (SSD) of Derosal data for the criterion group 4 soil macro chronic



Source: own illustration, darwin statistics & gaiac

Red line: fit of the log-normal distribution to the original data; blue lines: non-parametric bootstrap sampling (1000 runs) simulated uncertainty of the fit; dashed black line: 95 % confidence interval of bootstrap samples. In the databases, the Plant protection product name Derosal was used as a surrogate for the active ingredient carbendazim.

6.6 Comparison of suggested toxicity thresholds with criteria from earlier studies and proposals

After the herewith derived thresholds were applied and candidate lists of potentially classified substances and their assigned hazard classes were derived, those were further analyzed for consistency and relevance and were compared to existing proposals. To evaluate the threshold values of the different criterion groups derived in this chapter, these were compared with soil toxicity criteria formerly proposed in the literature and the final proposal at the UN level in 2006, which failed due to political reasons and cost-benefit considerations (see chapter 1). At that time, the available data basis was more limited. Compared to this, the analyses of the PROSOIL project were based on a much larger data basis across different regulatory areas.

The different historical approaches (rational and statistical methods) used to derive soil toxicity thresholds are briefly discussed in the following. In most of the listed literature, no detailed information was given on the data sources, quality and quantity, and the regulatory context. Most studies did not focus on a specific regulatory context such as CLP or PBT.

For Carbonell et al. (1997), no detailed information on the data sources was available. They based their derivation of acute and chronic terrestrial threshold values on different exposure routes, including uptake from the soil, oral exposure, uptake through inhalation, exposure via spray deposition, and above-ground organisms (Carbonell et al., 1997, Renaud et al., 2004).

The suggestions of Torstensson et al. (1999) were based on acute soil data from the ECHA IUCLID database (Torstensson et al., 1999, Renaud et al., 2004). However, the data quality and extent remained unclear, and information on how these data found entry into the derivation of thresholds was not given. The proposed threshold for an acute toxicity category 1 (EC_{50}) of 10 mg ai/kg sdw assumed that highly contaminated soils could reach contamination levels of up to 10 mg ai / kg sdw through atmospheric deposition and application of sewage sludge disposal (Vega et al. 2003, Torstensson et al. 1999).

More information was given in Renaud et al. (2004). The authors integrated the already mentioned approaches from Carbonell et al. (1997) and Torstensson et al. (1999) into their approach. This new approach was based upon data analyses of several regulatory databases (USEPA TERRETOX, PESTS 2000, ECHA IUCLID database, INRA AGRITOX). It resulted in an internally built database of acute earthworm toxicity data for 313 different substances. Within these substances, the majority were agricultural plant protection products (herbicides 36.1 %, fungicides 24.6 %, insecticides 14.4 %, acaricides or nematicides 5.4 %; mixed modes of action 6.1 %). Furthermore, Renaud et al. (2004) used the geomean and the quantile approaches, which were also used in the present study.

The final proposal at the UN level (UN 2006) and the proposal of the New Zealand EPA (NZEPA 2012) provided both acute and chronic threshold values to be used for the classification and labeling of substances. Unfortunately, information about the data sources, data quality, or derivation procedure were not included in the final publicly available documents as these were proposals for implementation in the legal text. The UN proposal was based on the literature, analyses, and studies mentioned above and developed by different member states and the OECD expert group. The confidential documents from expert groups at the EU and UN levels were not available for comparison.

Most of the former proposals only referred to criteria corresponding to a “very toxic” classification. The final UN 2006 proposal laid down “very toxic”, “toxic”, and “harmful” criteria and a safety net category (4) for different taxonomic/exposure groups.

Table 17 shows the individual threshold values per criterion group developed in this project. The table is further divided into the three approaches, geomean, quantile, and nhst. Finally, the threshold values, i.e., soil toxicity criteria, proposed in the past as cited above are listed in the table (Table 17).

The historical attempts to develop toxicity criteria focused on specific organisms and exposure routes, such as earthworms (Renaud et al., 2004), foliar or soil exposure, and bees or mammals. These were not entirely congruent with the criterion groups derived in this project where in-soil organisms were focused. Most of the cited publications did not give many insights into their procedures, data sources, and data handling. It often remained unclear why the authors of the studies focused on specific areas of soil toxicity. However, the assumption is evident that as the available data was not particularly exhaustive, the authors may have focused on specific data sets available in their area of interest. Most historical threshold suggestions are available for soil macroorganisms and the statistical endpoint EC_{50}. For the statistical endpoint NOEC, suggestions for soil-associated plants are available (UN 2006). Indeed, for plants, chronic values are used.

Table 17: Threshold values (soil toxicity criteria) proposed by historical approaches compared with individual threshold values [mg ai/kg sdw) per criterion group developed in this

project (divided into three approaches geomean, quantile, nhst). Data marked with an asterisk only cover earthworm data, not soil-macroorganisms.

criterion group	Group 1 plant acute	Group 2 plant chronic	Group 3 soil macro acute	Group 4 soil macro chronic	Group 5 soil micro chronic
VERY TOXIC					
UN 2006	10	1	10	1	1
NZEPA (2012)			1		
Renaud et al. (2004) Geomean			4*		
Renaud et al. (2004) Quantile			60*		
Torstenssen et al. (1999)			10		
Carbonell et al. (1997)		1	1		1
PROSOIL geomean threshold	0.2	0.1	5.5	2.6	0.2
PROSOIL quantile threshold	0	0	0.2	0.3	0.1
PROSOIL nhst threshold	0.1	0.1	0.4	0.4	0.3
TOXIC					
UN 2006	100	10	100	10	10
NZEPA (2012)			10		
Torstenssen et al. (1999)			100		
Carbonell et al. (1997)		10	10		1
PROSOIL geomean threshold	1.6	0.8	55.3	25.9	2
PROSOIL quantile threshold	0	0	0.4	0.5	0.1
PROSOIL nhst threshold	0.2	0.1	0.8	0.8	0.7
HARMFUL					
UN 2006	1000	100	1000	100	100
NZEPA (2012)			100		
Carbonell et al. (1997)		100	100		
PROSOIL geomean threshold	16.2	8.1	552.7	258.9	19.7
PROSOIL quantile threshold	0	0	0.8	1	0.2
PROSOIL nhst threshold	0.4	0.3	1.5	1.5	1.3

As shown in Table 17, the threshold values proposed in this project differ depending on the applied approach (i.e., quantile, geomean, and nhst approach) and the selected criterion group. Due to the differences in data basis and attribution to an organism group, the formerly developed threshold values can only be approximately compared to the values derived in this project. Furthermore, the historical datasets were assumed to hold less intrinsically toxic pesticides and more moderately toxic substances. However, the criterion groups of this project correspond to the ones used in former approaches: "Group 3: soil macro acute" corresponds to six historical values, whereas "Group 1: plant-acute", "Group 2: plant-chronic", and "Group 4: soil macro chronic" corresponds to only one historical proposed threshold each.

The correlation between the derived threshold values is not directly evident for all criterion groups and is prone to strong fluctuations between the approaches. But in the case of the criterion group "Group 3: **soil macro acute**" the range between the threshold values derived in this project is quite similar to the range throughout the historical approaches. The threshold values given by the historical approaches range from 1 to 60 mg ai/kg sdw for "very toxic", whereas the range of the threshold values derived in this project ranges from 0.188 to 5.527 mg ai/kg sdw – both approximately featuring a range of factor 30 to 50. For all other criterion groups, only one historical value is available. In general, it can be stated that only the geomean approach derived in this project provides higher values for the criterion groups 3 and 4 (soil macro). For all other values, the historical approaches used higher values. It remained unclear why the quantile and geomean approach from Renaud et al. 2004 did not correlate with the quantile and geomean approach – however, it is at least in the same range. Renaud et al. 2004, for example, only used data on earthworms and a smaller database.

It has to be clarified that the PROSOIL project aimed to derive toxicity thresholds on a comprehensive data basis and with different statistical methods to get a picture of the sensitivity of soil organisms and build a basis for decision-making on the best-suited approach (and should not anticipate this decision before discussions in an international group of stakeholders would have taken place). The discussion may then be the basis for setting criteria at the legal level. The UN (2006) proposal (and NZ EPA) already included the final criteria set based on scientific analyses in the legal text. Hence, a direct comparison was not meaningful, just a comparison if and in which cases PROSOIL came up with similar ranges.

6.7 Selection of the approach to derive toxicity thresholds

In principle, the three approaches tried in the present research were not equivalent but were based on different assumptions and underlying hypotheses (methods described in sections 6.1.3-6.1.5). The *quantile* approach maintains a general protection level by flagging a specific proportion of the substances as toxic, and the *geomean* approach classifies substances based on the fact that they are more toxic than the average. The latter approaches aggregated the data for each substance first and used the most sensitive ecotoxicological value only. Essentially, they perform rankings from different origins. In other ways, the NHST approach made use of all data available and thus could give an estimate of the probability of a substance being more toxic than the whole population of all available substances.

From this, it becomes apparent that the three approaches to deriving toxicity thresholds resulted in quite different threshold values and numbers of classified substances. Hence, the authors had to decide which approach would be finally used. At this point of the analysis, it was decided not to select one approach exclusively but to prefer the geomean approach in subsequent analyses of the relevance and for comparisons with historical approaches (chapter 7) whenever a choice had to be made for the sake of clarity and communicability. The outcome in terms of the total number of classified substances and thus the level of protectiveness for soil organisms did not play a decisive role. The results of this approach were compared to the historical approaches in the previous chapter 6.6.

6.8 Conclusion on the development of soil toxicity thresholds

According to the authors, even if data would tell their own story, the most sensible point for the derivation of toxicity thresholds depended on informed expert decisions of the assessors in the areas of conflict between principles of precaution and pragmatism. For example, the present study's authors used the minimum ecotoxicological value within each criterion group and substance to derive the "hazardous quantile" and the geometric mean. Furthermore, the authors did not correct for outliers, neither on the lower or higher tails of the distributions. Elimination of outliers would lead to shorter candidate lists using the quantile approach because the comprehensive list of substances would be shorter (the lower 10th quantile of 100 values comprises approx. 10 substances, the 10th quantile of 200 values comprises approx. 20 substances). Consequently, fewer substances would be deemed toxic. Formal ecotoxicity outliers (i.e., high values) could be technically eliminated but are believed to belong to the actual distribution of the datasets in principle. They represent substances of very low toxicity that were potentially tested at the limits of, e.g., water solubility.

Differences in toxicity between different assessment areas, which were filtered to uniform *ecotoxunits,* lead to variable weights of the areas when it comes to the derivation of toxicity thresholds. Studies using non-target arthropods, birds, and mammals were excluded from the analyses because they do not belong to the groups of in-soil organisms in a narrower sense, and the guidance documents do not cover them for soil organisms from different regulatory areas.

Three approaches were chosen from the literature and preceding scientific and regulatory initiatives to derive soil toxicity criteria: the "geomean approach", the "quantile approach", and the NHST-approach. The final value that is used as the threshold for hazard classification that is applied to a given list of chemical substances depends not only on the distinct approach but on specific criteria, such as the definite quantile chosen (quantile approach) and the safety factors used to extrapolate from toxic to very toxic or harmful substances (geomean approach). These criteria must be subject to intense discussion. Most sensitive species could differ between taxonomic groups and chemical classes and are not always standard test species.

7 Relevance of soil toxicity criteria

7.1 Comparison of suggested threshold values with existing classifications

It is critical to evaluate whether substances that are falling under the derived thresholds values of this project are already classified as "hazardous to the aquatic environment" within the existing class under CLP/GHS (with H-phrases corresponding to very toxic, toxic, or harmful to the environment), or are already identified as PBT, SVHC, or POP substances and thereby regulated to assess the relevance and benefits of the toxicity threshold values derived in this project. Therefore, the derived candidate lists of substances (based upon the threshold values developed in this study) were compared with the current ECHA C&L inventory to estimate how many substances from this list are already covered by the current (aquatic-based) classification. Furthermore, the candidate lists were compared with the substances' current PBT-, SVHC, and POP–status.

For the comparison with the C&L-inventory, a complete version was requested from ECHA in the form of an Excel spreadsheet (version from 12th October 2021). From this entire C&L inventory, both officially harmonized and self-classified substances ("notified") were examined. For this purpose, in particular, the columns "*CON C&L Harmonised – Context indicating whether there is a harmonized C&L agreed for the substance*" and "*CON C&L Notified – Context indicating whether the substance has been notified under the CLP Regulation*" were used as indicators, and the entries were considered separately between "harmonized" and "notified". Harmonized data were then filtered by the entries (H-phrases) listed in column "*CLH Classification Haz class & cat codes – If there are harmonized C&L entries for the substance, the hazard class and category code(s) are given*". Accordingly, for the notified data, column "*All C&L > Classification with % – From all C&L data available, the hazard class, category, and statement codes are aggregated by % of data submitters who indicated that class, category, and statement. E.g., |Acute Tox. 3|-|H311| (78.21%) means that 78.21% of all C&L notifications for a substance indicated Acute Tox. 3 in combination with H311*" was filtered by the entries (H-phrases). PROSOIL candidate substances not listed in one of these columns with a corresponding H-phrase were considered unclassified accordingly.

Table 18 shows the matching of the substances within the PROSOIL criterion groups with the environmentally relevant H-phrases of the current ECHA CL-inventory, i.e., based upon the complete candidate lists (including the derived hazard classes very toxic, toxic, and harmful) and divided into the three main statistical approaches, i.e., the geomean, the quantile and the nhst approach. This matching with H-phrases according to CLP shall provide an approximation of how many potentially toxic substances for soil organisms have already been classified hazardous for the aquatic environment. The acute-based H-phrases H401 (toxic) and H402 (harmful) have not been utilized in the present comparison, as they are only included in the GHS classification system but were not implemented in the European CLP system for the aquatic environment.

Table 18: Percentage of the PROSOIL candidate lists matching with the H-phrases of the ECHA C&L-Inventory, reflecting the % of substances already covered by the existing CLP system.

The hazard classes of the PROSOIL project, i.e., very toxic, toxic, and harmful, are compared with the corresponding H-phrase, plus, in case of toxic and harmful classifications, the respective higher classified H-phrase (marked with a "+"). The comparison was made for all three statistical approaches geomean, quantile, and nhst. Complete C&L-inventory data have been used (harmonized + notified entries) but separated between acute and chronic toxicity. The total coverage between C&L-inventory and PROSOIL candidate lists (independent of the hazard classes) is given together with the number of substances the values are based on (n). The whole analysis is based upon an ECHA C&L-inventory dump from 12th October 2021.

approach	group	hazard class	H400 [%]	H410 [%]	H411+ [%]	H412+ [%]	Total [%]	n
GEOMEAN	plant acute	very toxic	68.4				63.3	193
GEOMEAN	plant chronic	very toxic		64.9			78.5	442
GEOMEAN	plant chronic	toxic			74.6			
GEOMEAN	plant chronic	harmful				19.5		
GEOMEAN	soil macro acute	very toxic	70.9				57.3	398
GEOMEAN	soil macro chronic	very toxic		63.4			69.1	417
GEOMEAN	soil macro chronic	toxic			61.2			
GEOMEAN	soil macro chronic	harmful				64.2		
GEOMEAN	soil micro chronic	very toxic		81.8			87.7	81
GEOMEAN	soil micro chronic	toxic			90.0			
GEOMEAN	soil micro chronic	harmful				83.3		
QUANTILE	plant acute	very toxic	79.2				86.8	38
QUANTILE	plant chronic	very toxic		78.0			77.4	84
QUANTILE	plant chronic	toxic			57.9			
QUANTILE	plant chronic	harmful				60.0		
QUANTILE	soil macro acute	very toxic	90.0				70.3	64
QUANTILE	soil macro chronic	very toxic		87.5			79.5	83
QUANTILE	soil macro chronic	toxic			65.2			
QUANTILE	soil macro chronic	harmful				60.8		
QUANTILE	soil micro chronic	very toxic		87.5			81.8	11
QUANTILE	soil micro chronic	toxic			50.0			
QUANTILE	soil micro chronic	harmful				100.0		

approach	group	hazard class	H400 [%]	H410 [%]	H411+ [%]	H412+ [%]	Total [%]	n
NHST	plant acute	very toxic	69.2				67.2	128
NHST	plant chronic	very toxic		65.2			80.7	269
NHST	plant chronic	toxic			82.6			
NHST	plant chronic	harmful				84.6		
NHST	soil macro acute	very toxic	68.6				72.2	79
NHST	soil macro chronic	very toxic		83.4			81.0	95
NHST	soil macro chronic	toxic			62.5			
NHST	soil macro chronic	harmful				68.9		
NHST	soil micro chronic	very toxic		85.7			89.5	48
NHST	soil micro chronic	toxic			90.0			
NHST	soil micro chronic	harmful				92.9		

The different hazard classes within the candidate lists of the PROSOIL project, i.e., very toxic, toxic, and harmful, have been compared with the corresponding H-phrase plus the respective higher classified H-phrases to calculate a percentage matching between the C&L-inventory and PROSOIL. Thus, the substances within a PROSOIL criterion group classified as, e.g., harmful (to soil organisms), have been compared with the substances marked harmful, but also toxic and very toxic in the C&L-inventory, as also these would fulfill the requirement for "at least harmful". Thus, the given percentage values in Table 18 can be seen as a best-case coverage and provide information on PROSOIL substances classified equally or even higher within the current aquatic-based CLP assessment. These comparative analyses were strictly separated between acute and chronic classifications (e.g., group 3: soil macro acute has only been compared with H400, not H410/H411/H412).

Under REACH, a large part of substances is only self-classified ("notified") by companies registering the substance (registrants), while all classifications of active substances, approved as PPP or biocides, must be harmonized between the authorities. In the present analysis, the given data have been calculated for all C&L-inventory entries, i.e., harmonized data plus notified data (self-classified). For notified data, the percentage of data registrants who indicated a specific H-phrase was ignored to create a most conservative scenario (example: even if only one out of ten registrants have notified an H400 classification for a specific substance, this was regarded as an H400 substance in the present analysis). Furthermore, for notified data, the highest hazard class was decisive for the comparison (i.e., H410 > H411 > H412). A substance with multiple classifications by different submitters, but for example, at least one notified H410 classification, was excluded from comparison with lower classifications such as H411/H412. Thus, no substance was included multiple times in the total values. The total values given in Table 18 provide information on the total matching of a PROSOIL group (whole candidate list) with the C&L-inventory, independent of its hazard classification or H-phrase (i.e., any matching between candidate list and C&L-inventory). Thus, even for comparing acute classifications (i.e., PROSOIL very toxic vs. H400), the total value might be higher or lower than the pair-wise comparison since it includes the whole PROSOIL candidate list, not only the "very toxic" classification. The total value is a kind of total balance, which clearly shows to what extent any aquatic classification covers an entire candidate list.

In general, the number of substances per criterion group is essential for the reliability of the correlation. The number of substances assessed per group can be seen in Table 18. Groups with data for many substances, for example, plant chronic or soil macro chronic, can be regarded as statistically more reliable. For criterion groups with only a few different substances, such as soil micro chronic data, the comparison with existing hazard classifications should be considered cautiously.

Besides comparing the PROSOIL candidate lists with the ECHA C&L inventory separated between acute and chronic classifications (see Table 18), a second analysis has been done completely independently from the specific classification. For this analysis, all substances in a candidate list of a specific PROSOIL approach (geomean, quantile, NHST) have been checked for their classification with the H-phrases H400, H410, H411, H412, or H413, while completely ignoring acute/chronic labeling and very toxic/toxic/harmful classifications. The results of this superordinate analysis can be seen in Table 19.

Table 19: Superordinate analysis of the percentage of PROSOIL candidate lists matching with any environmental H-phrase of the ECHA C&L-Inventory, reflecting the total percentage of substances covered by the existing CLP regulation. The analysis has been performed entirely independent of any acute/chronic labeling or very toxic/toxic/harmful classification while checking the matching of the substances with all H-phrases H400, H410, H411, H412, or H413 and for all three statistical approaches geomean, quantile, and nhst. Harmonized, notified, and complete C&L-inventory data have been used. The whole analysis is based upon an ECHA C&L-inventory dump from 12th October 2021.

Approach	Group	C&L harmonized entries	C&L notified entries	C&L all entries
GEOMEAN	group 1plant acute	53.4	31.1	**84.5**
GEOMEAN	group 2: plant chronic	43.0	37.6	**80.5**
GEOMEAN	group 3: soil macro acute	44.7	32.7	**77.4**
GEOMEAN	group 4: soil macro chronic	37.6	34.1	**71.7**
GEOMEAN	group 5: soil micro chronic	56.8	32.1	**88.9**
QUANTILE	group 1: plant acute	55.3	39.5	**94.7**
QUANTILE	group 2: plant chronic	40.5	39.3	**79.8**
QUANTILE	group 3: soil macro acute	60.9	25.0	**85.9**
QUANTILE	group 4: soil macro chronic	54.2	28.9	**83.1**
QUANTILE	group 5: soil micro chronic	18.2	63.6	**81.8**
NHST	group 1: plant acute	54.7	32.0	**86.7**
NHST	group 2: plant chronic	46.1	37.2	**83.3**
NHST	group 3: soil macro acute	60.8	27.8	**88.6**
NHST	group 4: soil macro chronic	54.7	29.5	**84.2**
NHST	group 5: soil micro chronic	56.3	33.3	**89.6**

The matching of the PROSOIL candidate lists with the aquatic classification contained in the ECHA C&L-inventory differs strongly between the PROSOIL criterion groups. The total percentage of matching entries provides the most reliable average measure, thus, independent of the specific classification or H-phrase but based upon a separation between acute and chronic data. Here, the PROSOIL candidate lists reach a total matching with the C&L-inventory between 57.3 % (GEOMEAN approach, group 3: soil macro acute) and 89.5 % (NHST approach, group 5: soil micro chronic), reflecting the % of substances already considered by the existing (aquatic) classification.

The situation is similar for the highest toxicity classification, “very toxic”, and the respective comparisons with H400 resp. H410. Best matching of 90 % occurs in Group 3: soil macro acute (QUANTILE approach) but decreases to 63.4 % for group 4: soil macro chronic (GEOMEAN approach).

Apart from minor variations, the percentage values differ relatively little between acute and chronic groups. However, the high number of substances included in group 2: plant chronic and group 4: soil macro chronic, seems to be the most dedicated group for further analyses. It can be concluded that far more than half of the candidate chemicals across all PROSOIL criterion groups, i.e., most substances, are already covered by aquatic classification. However, an average of 20 to 30 % of the substances regarded "very toxic", "toxic", or "harmful" for the different groups by the PROSOIL approaches are still not covered by the ECHA C&L inventory and thus, might not be sufficiently classified to protect soil-organisms optimally.

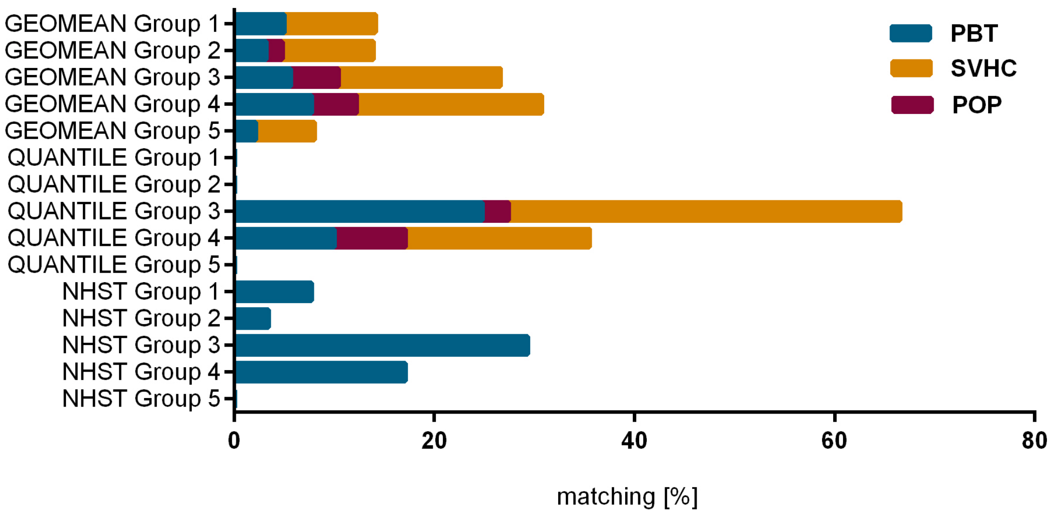
The situation is not entirely different but also refers to the superordinate total percentage of matching entries, thus, independent of any classification or labeling (i.e., no separation between acute and chronic data). Here, matching is higher but still in the range of approximately 70 to 90 % (harmonized + notified entries), i.e., 10 to 30 % of the candidate substances not classified within the current CLP regulation.

It is to be considered that the analysis was done for the separate criterion groups. In practice, classification is done more or less "across" the taxonomic groups – only separated for acute and chronic hazards. The most sensitive endpoint available is decisive for the classification as "hazardous to the aquatic environment with long-lasting effects". It was not further specified which organism or test the classification was based on (although criteria are defined for distinct taxonomic groups).

7.2 Comparison of suggested threshold values with existing PBT, SVHC, or POP-status

Furthermore, the percentage matching of the PROSOIL criterion groups with current PBT, SVHC, and POP listed substances (all of which are related to aquatic studies only when considering environmental criteria) is examined. The percentage matching of the complete candidate lists with PBT, SVHC, and POP status is given in Figure 18 and ranges between 0% and around 40% (a maximum matching of approximately 40 % can be seen for group 3: soil macro acute (quantile approach) with the SVHC lists). A minority of 10 to 20 % of the substances listed in the PROSOIL criterion groups are identified as PBT, POP, or SVHC.

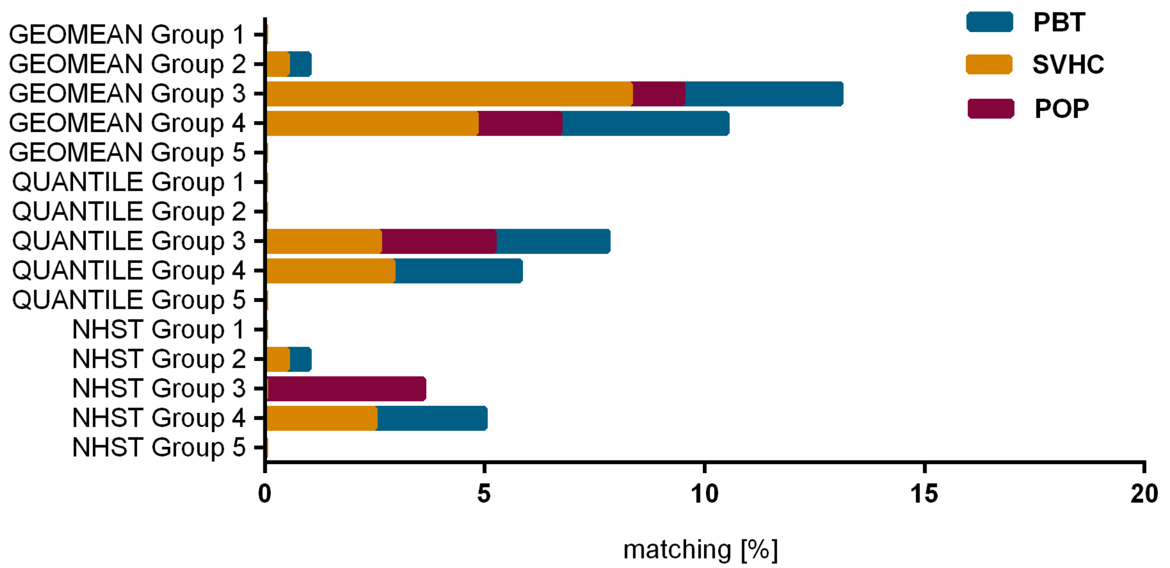
Figure 18: Percentage matching of the PROSOIL criterion groups (group 1 = plant acute EC50, group 2 = plant chronic NOEC EC10, group 3 = soil macro acute EC50, group 4 = soil macro chronic NOEC EC10 add EC50, group 5 = soil micro chronic NOEC EC10) for three statistical approaches with current PBT, SVHC, and POP listed substances.



Source: own illustration, darwin statistics & gaiac

The situation looks different if only the substances classified “very toxic” within the PROSOIL criterion groups are matched against current PBT, SVHC, and POP lists (Figure 19). SVHC still makes up the most considerable portion but shows only a maximum of approximately 15 % for group 3: soil macro acute (geomean approach). All other matching values are below 5 % per group. These low values are surprising since a higher coverage should be observed for the substances classified as “very toxic” rather than the entire candidate lists. It can be observed that substances of very high concern (SVHC) do not primarily occur as “very toxic” for soil organisms but within the toxic or harmful hazard classes.

Figure 19: Percentage matching of the substances classified “very toxic” within the PROSOIL criterion groups (group 1 = plant acute EC50, group 2 = plant chronic NOEC EC10, group 3 = soil macro acute EC50, group 4 = soil macro chronic NOEC EC10 add EC50, group 5 = soil micro chronic NOEC EC10) for all three statistical approaches with current PBT, SVHC, and POP listed substances.



Source: own illustration, darwin statistics & gaiac

7.3 Conclusions on the relevance of soil toxicity thresholds & toxicity criteria

The percentage values from comparing the different PROSOIL criterion groups and the C&L inventory, PBT, SVHC, and POP lists are similar. The similarity of percentages could lead to the assumption that the criterion groups differ only slightly in their substance compositions. Thus, before a general conclusion can be drawn, the differences between the various groups and candidate lists must be examined in more detail.

A deeper look at the individual candidate lists shows that the groups differ enormously in their chemical composition in most cases. A dissimilarity analysis has been conducted, stepwise comparing the composition of substances of all PROSOIL criterion groups, and resulting in a percentage of dissimilarity (Figure 20) that confirmed the significant dissimilarities between the candidate lists. The minimum dissimilarity was observed between group 2: plant chronic and group 1: plant acute, amounting to approximately 60 %, reflected by a similarity between the two groups of about 40 %. All other groups differ even more in their composition of substances.

So, the comparison between the PROSOIL groups and the existing aquatic classifications is not only based on one similar selection of substances across the groups but on five very different compositions of substances (i.e., the criterion groups).

Figure 20: Dissimilarity analysis of the PROSOIL criterion groups. The y-axis describes the dissimilarity between the substances listed in two pairwise compared groups (0 = equal substances; 1 = not one single similar substance).

Source: own illustration, darwin statistics & gaiac

Keeping in mind the very different compositions of the PROSOIL criterion groups, the percentage values of the individual groups gain weight. The general matching of the criterion groups with the C&L inventory, PBT, SVHC, and POP lists seems plausible. It proves that the aquatic assessment is not generally appropriate to protect soil organisms regardless of the statistical approach adequately. In turn, arguments are provided for the additional implementation of complementary soil toxicity criteria within the existing "Hazardous to the environment" category. However, the evaluation shown here is only a first attempt.

8 Stakeholder involvement: The International PROSOIL workshop

Stakeholders were involved in integrating more diverse viewpoints, embracing historical experiences, gaining acceptance of the concept developed so far, and projecting future activities towards implementation into international legislation. The workshop was jointly organized, moderated, and reported by darwin statistics, gaiac, and UBA and hosted 55 international experts in soil ecotoxicology and chemical regulation. The workshop was the final part of the PROSOIL project. In the following, the presentations and primary outcomes of the discussions are summarized. The reported findings correspond to the summary and workshop documentation were distributed to the participants after the Workshop and made available to UBA. The results of the Workshop are further discussed in chapter 9.

Session I: Background and Motivation
Background on historical and recent initiatives to develop toxicity criteria for terrestrial organisms in the context of classification and labeling and PBT assessment (presentation and plenary discussion)

Ecotoxicity assessments in CLP and PBT are based solely on aquatic toxicity criteria. Around 15 years ago, **historical initiatives** and proposals to establish terrestrial toxicity criteria for classification and labeling at the EU and UN levels did not reach a consensus (European Commission (2001). They led to a proposal at the UN level in 2006 (Committee of Experts on the Transport of Dangerous Goods and the Globally Harmonized System of Classification and Labelling of Chemicals). However, initiatives were stopped in 2008 due to political reasons, cost-benefit considerations, and lack of data. In 2020, the EU Commission announced in their "Chemical Strategy for Sustainability" to check the feasibility of including terrestrial toxicity criteria into the CLP regulation again.

The **PROSOIL project** was building on literature research of publicly available proposals, analyses, and regulatory data and aims at the compilation of a database on soil ecotoxicity data, the identification of toxicity ranges for different soil organism groups to propose toxicity criteria for use in the context of classification and labeling (CLP) and PBT assessments. The project focused on terrestrial organism groups potentially exposed to toxic substances via soil, i.e., plants, invertebrates (e.g., oligochaetes, collembolans, mites), and microorganisms. Above-ground organisms such as bees, birds, or mammals are not considered. A comprehensive set of soil ecotoxicity data for chemicals across different chemical products' legislation (REACH, plant protection products, biocides, and pharmaceuticals) were considered, with varying data availability, quality, and coverage due to different data requirements.

This **workshop** aimed to put the issue on the agenda again and discuss options and data-need to derive toxicity criteria for soil organisms based on a new comprehensive database, analyses, and proposals elaborated during the PROSOIL project.

In preparation for the workshop, participants were asked to fill in a *short questionnaire* on their **motivation, personal expertise, and viewpoints**, i.e., data quality, toxicity thresholds, and the need for additional protection for soil organisms. The results showed that most participants were experienced in environmental assessments focusing on the soil compartment. While the current protection level for in-soil organisms was perceived to be better under the plant protection product regulation than under CLP and REACH, most experts agreed that the inclusion of terrestrial criteria for CLP and PBT would be beneficial and a future "soil framework directive" would be desirable. It was agreed that the derivation of threshold values should be as transparent as possible, that diverse data sources could be considered suitable for an overall generic assessment, and that data availability was considered crucial. Chronic and more complex ecotoxicological test data is still comparatively poorly available.

Session II: The PROSOIL database

Overview of the PROSOIL database: data quality and availability, groups defined for analyses (presentation and plenary discussion)

For the compilation of the PROSOIL database, a **comprehensive set of ecotoxicity data** (around 125.000 entries) for various organisms (plants, in-soil microorganisms, in-soil macro-organisms), retrieved from different regulatory repositories (UBA ICS [38.000 data entries before harmonization, quality checks, and duplicate removal], ECHA IUCLID [43.000], UBA ETOX [3.000] and US EPA's ECOTOX [131.000]) was compiled. Data were harmonized concerning unified species identifiers, endpoint nomenclature, and test concentration units. Non-plausible outliers were discarded to have a reliable distribution of ecotoxicity values. About 50 % of the data entries were not useable due to missing entries for unique identifiers (CAS numbers), test species, ecotoxicological values, or duplication. Most data entries were retrieved for plants and lumbricids, with plant protection products (active substances) reported as the dominant substance class.

Most data were retrieved from the US EPA ECOTOX, particularly for plants and chronic data for soil macro-organisms. During the plenary discussion, a representative of DG Environment described the **current plans of the EU Commission** to check the inclusion of terrestrial criteria into the current CLP and UN GHS classification (e.g., harmful to the aquatic **and** terrestrial environment). A first proposal was promised for 2022, followed by discussions and agreements at the UN GHS level. Ongoing activities on the inclusion of CLP criteria for PBT and PMT were also mentioned. Three experts involved in the historic initiatives at the EU and UN levels to establish soil toxicity criteria for CLP/GHS described the main challenges. Data availability and quality and the importance of demonstrating the benefits were crucial at that time. It was agreed that the workshop's conclusions constitute an essential trigger for further discussions and could support the development of proposals and decision-making processes.

The dataset analyzed 15 years ago focused on around 600 entries for chemicals. While the participants highly appreciated the comprehensive data set compiled here, the **challenges regarding data quality** were raised as an essential issue to get broad support for a new criterion. Many of the data retrieved had not undergone validity checks by regulatory authorities and often comprised non-standard studies from peer-reviewed scientific literature (US EPA ECOTOX) or studies submitted by companies (ECHA IUCLID). On the contrary, some plant protection product data (UBA ETOX) comprised more tests and were checked in more detail. It was discussed how data quality could be best defined. In the light of the enormous dataset, the reliability of individual test results was checked by the processors of the respective databases following the strict rules laid down, while the PROSOIL project applied statistical methods to the data. As the database contained confidential data, e.g., on active substances for plant protection products, biocides, or pharmaceuticals, it will not be made publicly available.

Session III: Estimating plausible toxicity ranges

Statistical approaches to estimate threshold values, toxicity ranges for different terrestrial organism groups, and plausible toxicity criteria (presentation and plenary discussion)

Critical aspects of exploratory data analyses were presented to understand the data set and **distributions of ecotoxicity values** for different homogenous groups of soil organisms and decisions on the selection of groups. The distributions of ecotoxicity values were considered similar to pool data from different data sources and taxa. For example, the invertebrate taxa Collembola, Lumbricidae, Enchytraeidae, Nematoda, Isopoda, Myriapoda, and Acari were grouped as in-soil macro-organisms. Thereby, it was differentiated between acute and chronic data. Hence, five criterion groups were subjected to further analyses, threshold derivation, and definition of hazard categories: plants acute and chronic, macro-organisms acute and chronic, and microorganisms chronic. Acute data for macro-organisms was primarily available for earthworms, with *Eisenia fetida* as the most frequent test species, but it was often not the most sensitive one. While the UBA ETOX database held most entries for acute data, chronic data on soil macro-organisms was dominant in the US EPA ECOTOX database.

Different approaches were compared, i.e., the „geomean“, the „quantile“, and the „null-hypothesis significance testing (NHST)” approach to derive **ecotoxicity thresholds**. Based on the derived thresholds for each approach, the distinct categories harmful, toxic and very toxic were build (i.e., geomean: toxic = geomean, very toxic = geomean/10, harmful = geomean*10; Null Hypothesis Significance Test: toxic = significance threshold < 5% error probability, very toxic = significance threshold /2, harmful = significance threshold *2; quantile: toxic = lower 10th quantile, very toxic = quantile/2, harmful = quantile*2). The number of substances falling below the thresholds varied between the different statistical approaches. Statistical stability simulations (bootstrapping procedures) were conducted to check if the different approaches led to different conclusions when new data became available.

Plenary discussions were centered on a **better understanding and transparency of the reliability and comparability of data** as an essential basis for any analyses, the derivation of thresholds, and toxicity ranges. For the PROSOIL database, efforts were undertaken to harmonize the units of ecotoxicity endpoints to soil concentrations in mg active ingredient /kg dry weight of soil whenever possible. Moreover, it must be considered that data was compiled from diverse data sources and different test protocols within each of the four individual databases. A correction in the sense of normalization (e.g., by temperature) was also proposed. It was discussed whether the analysis of a more minor but well-known and checked sub-set of data could be beneficial or whether statistical approaches could be applied to improve or better understand data reliability. Further discussions are also needed concerning the **approaches used for deriving thresholds and hazard categories**. While the workshop presentations exemplified the results of the geomean approach, some participants pleaded for a probabilistic approach.

Session IV: Consequences, relevance, and benefits

Possible toxicity criteria and their “consequences”: relevance and benefits for a better consideration of terrestrial organisms (presentation and plenary discussion)

Toxicity ranges and threshold values (for the categories “very toxic”, “toxic,” and “harmful”) of the five criterion groups were compared for the results of the geomean, quantile, and NHST approaches. Substances were then compiled in **“candidate lists”** for each criterion group, i.e., listing those substances that would be classified as they were falling below the respective toxicity threshold. The differences between the **criterion groups** were evaluated concerning the identities of candidate substances. It was shown that the differences between the candidate substances of the data groups showed high pairwise dissimilarity indices for plants acute vs. chronic. The geomean approach would lead to the highest number of classified substances, possibly due to the higher threshold value applied to the other two approaches.

Further analyses were undertaken to get insight into the **most sensitive test species and affiliation to the regulatory area** of the candidate substances potentially classified. The most sensitive species was often *Eisenia fetida,* which might be due to its dominance in the source data as a standard test species over the last decades. The highest percentage of substances for all defined criterion groups came from plant protection products, followed by REACH or literature data (without any specified use), while pharmaceuticals and biocides were far less represented (also fewer data available in the database).

To discuss the **potential benefits and relevance of toxicity criteria** for soil organisms, the potential terrestrial classifications for candidate substances were compared with the existing aquatic classifications (H-phrases from the CLP inventory on self-notified or harmonized classifications). It was analyzed whether a candidate substance classified as harmful, toxic, or very toxic was already classified in the same or more critical aquatic hazard category for each criterion group and for harmonized- and self-classifications ("notifications") separately. Overall, these analyses indicated the benefits of introducing additional terrestrial toxicity criteria, as many substances from the candidate lists (10 - 30 %) relevant for the terrestrial environment were not yet classified by the existing regulations on ecotoxicity data for aquatic organisms. Moreover, it was also shown that only a low percentage of candidate substances were already identified as POPs, PBTs, vPvBs, and SVHCs. As for these assessments, the latter was evident that complex criteria (e.g., P, B, and T or ED properties) were applied, and the status is agreed upon on the European or international level (except for PBT screenings by companies).

In comparison to the terrestrial toxicity criteria of the historic proposal at the UN level in 2006, the thresholds derived by the PROSOIL project (geomean approach) were in similar orders of magnitude (with a factor of 2-5) for most of the criterion groups, except for plants (factor 10-50). In general, the threshold values tended to be consistently lower in this project (for the geomean approach given as units of "mg ai/kg sdw"):

- plant acute EC_{50} 0.2 - 1.6 - 16.2 (UN 2006: 10 – 100 - 1000)
- plant chronic NOEC 0.1 - 0.8 - 8.1 (UN 2006: 1 - 10 - 100)
- soil macro acute EC_{50} 5.5 - 55.3 - 552.7 (UN 2006: 10 – 100 - 1000)
- soil macro chronic NOEC 2.6 - 25.9 - 258.9 (UN 2006: 1 – 10 - 100)
- soil micro chronic EC_{50} 0.2 – 2.0 - 19.7 (UN 2006: 1 – 10 - 100)

During the plenary discussions, it was proposed to flag the **reliability of data entries** and information on the experimental conditions in future versions of the database and discuss biases, to strengthen confidence in the data basis. The need for improved availability and harmonization of data internationally was pointed out. It was considered essential to gather **arguments to demonstrate the potential benefits** of introducing complementary terrestrial toxicity criteria. The comparison with the existing aquatic hazard classification was emphasized as a measure, e.g., the possible number of substances that were not yet classified. A comparison of a similar analysis with the same approach used for deriving the aquatic thresholds was proposed. The classification system was meant to discriminate and highlight hazardous or toxic substances. However, not the sole number of additional classifications should be taken as an argument, but also that terrestrial criteria are relevant to protecting soil organisms. The main task of CLP as a hazard communication tool (e.g., for transport, occupational health, substance authorizations, and assessments) was highlighted, together with the consequences of new terrestrial criteria for the consideration of the soil compartment in future legislation.

Session V: Options, challenges, and further steps

Introduction, allocation to break-out groups, and discussion breakout groups.

In *breakout group discussions,* four topics were deepened by the participants: the assurance of data quality, the criterion groups defined, the approaches used for threshold derivation, and the relevance of terrestrial toxicity criteria.

Concerning the **data basis**, the main emphasis was put on the transparency and reliability of the data. The transparency was related to the flagging of the regulatory areas, data origins, and guideline conformity. A standardized system of reliability should be established, preferably after international consensus. Statistical tools could be of beneficial use for quality checks. In general, it was said that since extensive datasets are difficult to assess, the use of smaller, well-known and reliable datasets would be beneficial. Also, the regulatory affiliation of substances should be considered (i.e., intended use of plant protection products vs. unintended exposure of REACH chemicals). Also, remaining biases need to be checked and whether certain substances, such as plant protection products, drive toxicity thresholds.

For the defined **organism groups**, experts agreed that the groups "plants", "macroorganisms" (covering seven sub-groups as mentioned above), and "microorganisms" would be appropriate for use in sub-categories with terrestrial toxicity criteria. Other groups that might be considered are above-ground organisms, such as bees, mammals, or birds, which were not analyzed in the PROSOIL project due to the focus on organisms exposed via soil. Experts referred to differences in exposure routes during testing for different organisms or regulatory requirements, i.e., chemicals mixed into soil or sprayed onto test organisms or on soil, or exposures via air or pore water. Differences between intentionally applied plant protection products versus diffuse unintended exposures to chemicals were discussed. Concerning the approaches used, further discussion seems to be necessary.

Concerning the **relevance**, the protection of soil and the potential benefits of toxicity criteria were shared by the participants. Most participants supported a future introduction of terrestrial toxicity criteria. However, details need further discussion, such as the inclusion of terrestrial criteria into a new hazard class "hazardous to the terrestrial environment" or as an own hazard class in addition to the existing hazard class "hazardous to the environment". Terrestrial criteria were also seen as beneficial in the context of PBT assessments.

Overall conclusions and next steps

Current plans, further initiatives & next steps

The **PROSOIL database** was appreciated as a comprehensive dataset. The need for a good understanding of the origin and reliability of data was intensively discussed throughout all sessions. Most data were available for plant protection products and REACH chemicals. The detailed explorative analyses, harmonization exercises, and first proposals were appreciated.

However, the **approaches to derive toxicity thresholds and the derived criteria** need further discussion compared with the analyses and proposals of the historic initiatives at the UN level. The defined **main groups for categories** plants acute and chronic, macro-organisms acute and chronic, and microorganisms were agreed on.

In terms of the benefits, preliminary results indicate that many new hazardous substances to soil organisms were not yet classified by the existing aquatic criteria and were not identified as PBT- or POP substances or SVHC. Terrestrial criteria were considered beneficial to reflect better the soil compartment and associated consequences in further legislations relying on CLP/GHS.

The **discussions and feedback during this workshop** were highly appreciated and welcome afterward. The project aims at contributing to the ongoing discussions at the EU and UN levels for the development of terrestrial toxicity criteria. Further analyses of well-defined data sets and elaborations of more concrete proposals and analyses for above-ground organisms (bees, mammals) seem beneficial but beyond the PROSOIL project. Concerning the need for harmonized data across all regulations, the envisaged OECD data platform “Global Chemicals Knowledge Base” and the “One Substance One Assessment” initiative of the EU COM might be promising.

9 Conclusions

9.1 Relevance of explicit soil criteria for CLP and PBT

The main driver of the present research was the **protection of the populations and the biodiversity of soil organisms** from adverse effects that arose from chemical pressures in the environment. The project's initiation was based on past activities on how terrestrial and soil criteria should be integrated into hazard characterization and provisions for classification and labeling or PBT assessments. The protectiveness of the current criteria, which were only based on aquatic data, was analyzed. Hartmann et al. (2014) stated on a less comprehensible database that hazards derived from aquatic toxicity data generally delivered more protective classifications than for data from terrestrial compartments. However, exceptions were seen where toxicity was higher for the terrestrial system, and aquatic data could not cover the complete environmental hazards from chemicals (Renaud et al., 2004). It had to be concluded that specific soil-based toxicity criteria for classification purposes urgently needed to be discussed in a broader regulatory and scientific context. It should probably be implemented in European and international law wherever chemical legislation utilizes hazard-based criteria. Knowledge on toxicity thresholds of chemical substances for the environment may be very useful for prioritizations for further regulation. Thresholds could be established for classifications under CLP or GHS and PBT assessments. In addition, knowledge of toxicity thresholds can also be used for deriving specific soil toxicity criteria for contaminated sites where precautionary and limit values are needed to prevent humans and the environment from hazardous substances.

It was shown that across the three statistical approaches applied for the derivation of toxicity thresholds, 10 – 30 % of all substances within the *prosoildat* database across the five criterion groups were not classified by the aquatic criteria within CLP. The absolute proportion of classified substances depends on the concrete threshold values. However, none of the PROSOIL approaches led to a complete match with the existing classification. The differences did not exclusively occur at the sensitive end of the toxicity distribution at low values, but the sensitivity patterns of soil organisms' toxicity endpoints turned out to be different from the aquatic test organisms. This finding determines that no simple shift of the existing (aquatic) threshold or the new (soil) threshold would lead to complete coverage of the two classification approaches, but a separate or integrated assessment of soil organisms is necessary.

The lists of candidate substances in the highest hazard level of the developed soil toxicity criteria showed a low level of coincidence with existing exclusion criteria of the PBT, SVHC, or POP lists, i.e., that they are not yet fulfilling the respective criteria "persistency", "bioaccumulation", "toxicity", or "long-range transport". But you also must keep in mind that most of the above-mentioned hazard-based assessments have been still pending, and the absolute number of comparisons was relatively low. A more remarkable similarity was seen for the current CL inventory of ECHA on self-notified and harmonized classified substances (which were classified based on aquatic criteria), which was 70 – 90 % of the substances already classified by the existing system. Between 80 and 440 substances were classified at least "harmful" by the PROSOIL system, i.e., a relevant remainder would be newly classified if our proposal would be put into effect. These findings indicated that the exclusive use of the aquatic classification is not sufficient to rule out potential hazards for soil organisms. The PROSOIL-approach proved much more protective to soil organisms because more and other substances than derived by the aquatic toxicity criteria were classified and should thus be further developed and made ready for routine application shortly.

The toxicity thresholds suggested in chapter 6, derived by the three approaches, geomean, nhst, and quantile, were approximately in agreement (except for plants) with the threshold proposed by historical approaches concerning their absolute values and, accordingly, their protectiveness. However, the documents, sources from discussions at the EU and UN level, and the data basis leading to the UN proposal in 2006 were not publicly available and, therefore, difficult to follow and compare. The only source with a clearly stated data basis (Renaud et al. 2004) came for the criterion group "soil macro chronic" to a similar estimation of thresholds as in the present study. Consequently, from the definitions for soil toxicity criteria made herein, it was essential to have the same soil criteria for classification & labeling and PBT assessment of substances under various regulatory frameworks in the EU and worldwide. Consequently, more substances would be classified as harmful, toxic, or even very toxic, which is, in the end, a political decision. A higher level of protection for the crucial group of soil macro and microorganisms and the diverse functions they provide to productive and natural terrestrial ecosystems would be possible to implement in international law.

The relatively high number of substances that were not yet classified but would need to be labeled as hazardous to soil organisms to better inform on their hazards, resulting in a **low protection level of the recent legally binding procedures,** was a potentially alarming result of the research project (still depending on the selected thresholds). If confirmed, this would underpin the necessity of additional specific criteria for the soil compartment and should require attention and action from all stakeholder groups. Although the hazards and exposures of substances were already assessed during the environmental risk assessment under the different substance legislations (PPPR, BPR, REACH, pharmaceuticals), and their safe use is thereby to be also ensured for the soil compartment, classification and labeling would enhance the visibility and communication of their hazards to the soil.

Moreover, classification and labeling are triggering further measures, for example, in the context of chemical transport or workers' health. Moreover, it would require a consideration of these substances for further assessments, e.g., risk assessments for low volume chemicals under REACH. The PROSOIL project focused mainly on deriving thresholds directly translated into hazard classes for the CLP regulation.

Terrestrial criteria were also seen as beneficial in the context of PBT assessments. The T-criterion for PBT assessment could be either the "very toxic" class or extended to a lowered threshold using a specific conversion factor. Within PBT assessments, it would allow more data to fulfill the T-criterium, leading to a higher number of substances identified as PBT. And finally, the terrestrial compartment and associated data would be equally considered, which is currently primarily ignored under CLP and PBT assessments. Going beyond the CLP and PBT context, information on hazards to the soil compartment from the prospective assessments might also be necessary for the media-oriented frameworks to identify potential priority substances for environmental monitoring or other measures. Also, real-world monitoring data that helps in correlating pressures and biodiversity losses would be possible from large European-wide programs in collecting specimens and meta-data from the field and enhance the means of data-sharing by improving huge data-warehouses (Römbke et al., 2016). The monitoring data could then calibrate the C&L-system by actual chemical exposures in the environment.

9.2 Reliability of soil toxicity data

The **availability of soil-related toxicity data** was much higher than expected from the viewpoints of the scientists involved in project processing and the broader group of experts involved in the workshop discussions. The development of terrestrial cut-off criteria, e.g., by Carbonnell et al. in the year 1997, comprised approximately 1.000 chemical substances across all terrestrial assessment areas, including plants, birds & mammals, and earthworms. Which data was available for which group was not further specified? The *prosoildat* database held more than 3.700 different substances that covered a representative range of "intrinsic" physicochemical properties and the regulatory areas relevant to the CL of chemicals. The comprehensive database was considered a big step toward developing reliable, generalizable soil toxicity criteria. It was expected that the vast amount of available data from ecotoxicological studies would significantly increase in the following decades since the foreseeable innovation of guidance would come into force (e.g., by inventing new requirements in soil testing for plant protection products as by EFSA PPR, 2017). Overall, the **PROSOIL database** was regarded as a dataset of comprehensibility, with most data available for plant protection products and REACH chemicals.

Nonetheless, the database came with some characteristics and limitations that should be considered when interpreting the results. One of the most extensive data packages of the PROSOIL database originated from the US EPA ECOTOX system, which collected data from peer-reviewed literature (see sections 3.2, 4.3). This data supposedly lacked the formal validity and reproducibility criteria usually applied in a regulatory context where scientific experts at the authorities check the data in-depth within the approval processes of active ingredients, assuring high data quality and applicability for regulatory use based on internationally agreed criteria. On the other hand, the EPA data was checked during the peer-review process of the publisher by at least two or more independent experts in the field for the scientific excellence and the reliability of the (experimental and statistical) methodology for scientific purposes. The thorough, in-deep review process for scientific purposes and its broader scope in test species and chemicals made the open-literature data particularly useful for scientific exercises – allowing for data generated through "out-of-the-box" experiments. This characteristic was lacking the regulatory data.

Concerns were raised about the limitations of **transparency** since the primary data generating processes were not fully transparent and easily comprehensible. However, the approaches developed by PROSOIL, i.e., the data processing and all steps of the data analysis, were documented in detail by making the algorithms available via human-readable programming scripts. The most extensive part of the data was linked to the regulation of plant protection products which undoubtedly influenced the evaluation of the complete data set. The data requirements for the regulation of plant protection products included more data sets on organisms related to the compartment soil than, e.g., chemicals regulated within REACH or biocides. Therefore, the toxicity thresholds' derivation of terrestrial organisms within the evaluated data set might be influenced by organic chemicals used as plant protection products. However, this bias did not hamper the protectiveness of the overall evaluation of soil toxicity criteria.

The **usability of data** from various origins and a long history of developing soil ecotoxicological test protocols further hampers the analysis due to their enormous heterogeneity. However, no general semantics and guidelines providing ontologies and vocabularies of soil ecotoxicological data were ready-for-use. It was a vast and time-consuming step to make the four data sources interoperable by pre-defining common notations of endpoints, species, and guidelines. The harmonization processes assured that the PROSOIL project could impact the long term beyond the project's duration. The efforts that it would mean to implement soil toxicity criteria into hazard classifications seem much lower in terms of resources than the benefits of better protecting the ecologically outstandingly important group of soil organisms, which provide various supporting services in maintaining the productivity agricultural production systems worldwide.

9.3 Outlook & ongoing activities

According to Hartmann et al. (2014) and Renaud et al. (2004), most available information on the effects of chemicals on non-aquatic species has previously been focused on pesticides and veterinary medicines. Based on a more extensive dataset becoming available through REACH registration dossiers, further work could be amongst other considered examining if criteria based on aquatic organisms provide sufficient protection of the environment as a whole, including the non-aquatic organisms, or investigating appropriate hazard classification criteria for the terrestrial environment utilizing the availability of new data on non-aquatic organisms through the registration dossiers submitted under REACH. The PROSOIL project has already addressed these issues and might therefore be able to support the ongoing activities to include soil toxicity criteria for classification purposes or PBT assessments. However, some issues remain to be targeted to enhance the expressiveness of the project's outcome.

9.3.1 Needs concerning data analyses

The approaches to derive toxicity thresholds and the use of the derived criteria still need some discussion, especially in terms of reliability of data sources and comparison with the analyses and proposals of the historic initiatives at the UN level. After the involvement of an experienced group of experts, including those already engaged in the historical initiatives, there was a broad agreement that the next steps would be enhancing the transparency and reliability of the data rather than the integration of more and new data. Related to improved **reliability of the data**, future projects could include flags for each observation on the reliability and origin of the data point, the guideline conformity, and usability for meta-analyses. A standardized system of reliability assessments, such as KLIMISCH or CRED, which are standard for regulatory purposes, could be used even more extensively for terrestrial tests. The effort required to implement such an evaluation system would be enormous and should not be underestimated; however, it could address the existing concerns that, for example, large datasets available from public sources like the US EPA database are not quality-checked in a similar manner as the regulatory checked data for PPP from, e.g., the UBA ICS database. It cannot be conclusively determined whether a regulatory assessment generally increases the quality of the data and studies used compared to previous assessments (e.g., peer-review). However, the concerns could be addressed with further reliability assessments if feasible.

Statistical tools could be of use for quality checks in general. Reservation against extensive datasets exists; hence, for the sake of traceability and comparability, the use of smaller, well-described and reliable datasets would be considered beneficial for some analyses (i.e., when it comes to concrete regulatory measures). Also, the regulatory affiliation of substances could be considered (i.e., intended use of plant protection products vs. unintended exposure of REACH chemicals). Regardless of this, remaining biases need to be checked and whether certain substances, such as plant protection products, drive toxicity thresholds.

A definite possibility for further analyses would be conducting the analyses with the proposed UN criteria from 2006 using the extensive PROSOIL database and including the generation of candidate lists. Also, a large-scale analysis of aquatic data using the same procedures (analyses of toxicity distributions) as for soil toxicity data would be beneficial to have a trustworthy basis for comparison besides the ECHA C&L inventory and with the same "rational" background.

The PROSOIL project can contribute to the future implementation of soil toxicity criteria by demonstrating the range of toxicity distributions for selected organism groups. The next step could be to conduct more detailed analyses focusing on subsets of data for different regulatory areas to consider the underlying data quality and quantity better and discuss consequences.

For the defined **criterion groups**, it was concluded that “plants”, “macro-organisms,” and “microorganisms” would be appropriate for use in sub-categories of terrestrial toxicity criteria. These were agreed on in both the historic initiative and our proposal. Other groups that might be considered in the future are above-ground organisms, such as bees, mammals, or birds, which were not analyzed in the PROSOIL project due to the focus on organisms exposed via soil.

9.3.2 Needs for future implementation of soil toxicity criteria

The present research fits into the ongoing activities within the **EU policy-making** processes. The discussion during the international workshop showed that the European Commission is currently planning to develop strategies to integrate soil toxicity criteria in the upcoming revision of the CLP regulation, possibly as a comprehensive “terrestrial” or an overall T-criterion. Most stakeholders of the broader area of hazard-based assessments supported a future introduction of terrestrial toxicity criteria. However, details need further discussion, such as the inclusion of terrestrial criteria into a new hazard class “hazardous to the terrestrial environment” or as part of the existing hazard class “hazardous to the aquatic environment”, which would call for an adaptation of the hazard class to “hazardous to the environment”. PROSOIL approaches contributed to the ongoing policy-building processes while adding supporting information on the usability and analysis of large databases of soil ecotoxicity values. The ambition of the EC is even more prominent since the EU soil strategy recognized the outstanding meaning of healthy soils in an era of climate change and dramatic losses of biodiversity across all ecosystems, particularly for soils worldwide (European Commission 2021). This communication paper by the EU Commission also listed the other important and recent initiatives of the EU COM that were directly or indirectly related to soil protection (e.g., Farm-2-Fork strategy, Common Agricultural Policy, Zero Pollution Action Plan, Chemicals Strategy for Sustainability). Concerning the need for harmonized data across all regulations, the envisaged OECD data platform “Global Chemicals Knowledge Base” (https://www.echemportal.org/) and the “One Substance One Assessment” initiative of the EU COM (Van Dijk et al. 2021) might be promising. The description of the status of soil biodiversity and related threats and pressures towards soils are still widely ignored by actors from nature conservation and policy (Guerra et al., 2021), while experts' elicitation clearly showed that there was evidence that soils are threatened by various factors, amongst which the most relevant was direct overexploitation by humans and chemical pressures from, e.g., plant protection products (Orgiazzi et al., 2016). Maps that interpolate to the whole European Union depicted that most EU countries´ soils exhibit excess portions of highly vulnerable soils, with high risks for continuous losses of soil biodiversity.

Regardless of the information currently required for the hazard assessment of a chemical substance and in light of the precautionary principle underlying most European substance regulations, **all information available** (including intrinsic properties of a substance) should be taken into account that helps in, e.g., the identification of substances of very high concern (SVHC) (De Sadeleer, 2019). Indeed, some soil toxicity data are already used during specific SVHC identifications.

The PROSOIL project aimed to contribute to the ongoing discussions at the EU and UN levels to develop terrestrial toxicity criteria. Further analyses of well-defined data sets, analyses for above-ground organisms (non-target arthropods, bees, mammals), and elaborations of more concrete proposals are recommended for future activities based on the work at hand.

10 References

Aldenberg T, Jaworska JS (2000) Uncertainty of the Hazardous Concentration and Fraction Affected for Normal Species Sensitivity Distributions. Ecotoxicology and Environmental Safety 46:1-18

Alves PRL, Cardoso EJBN (2016) Overview of the Standard Methods for Soil Ecotoxicology Testing. Intech Open Science. Invertebrates - Experimental Models in Toxicity Screening. pp. 35-56; DOI: 10.5772/62228

Baath R (2014) Bayesian First Aid: A Package that Implements Bayesian Alternatives to the Classical *.test Functions in R. UseR! 2014 - the International R User Conference.

BBA - Federal Biological Research Centre for Agriculture and Forestry (2001) Auswirkungen von Pflanzenschutzmitteln auf die Reproduktion und das Wachstum von Eisenia fetida / Eisenia andrei. Richtlinien für die Prüfung von Pflanzenschutzmitteln im Zulassungsverfahren Teil IV 2 - 2, Januar 1994

Bosma TNP, Middeldorp PJM, Schraa G, Zehnder AJB (1997) Mass transfer limitation of biotransformation: Quantifying bioavailability. Environmental Science & Technology 31 (248-252).

Carbonell G, Aycart M, Callaba A, Escobar A, Fresno A, Pablos MV, Palma A, Ramos C, Santiago D, Tarazona JV, Vega MM (1997) Environmental hazard classification criteria for chemical substances. An integrated classification approach to identify the danger of chemical substances to terrestrial ecosystems. Development of specific criteria. Edited by the Spanish Ministry of Environment. (ECBI/78/95 Add 19)

Clausen H (1992) On the need for classification of chemicals for effects on the terrestrial environment In: Approaches for a hazard system for the terrestrial environment. Published by the European Chemcials Bureau, E.C.B. S.P.I. 99.80. pp 183-187

Committee of Experts on the Transport of Dangerous Goods and on the Globally Harmonized System of Classification and Labelling of Chemicals (2006) Environmental Hazards - Classification criteria for the terrestrial environment. UN/SCEGHS/12/INF.5

De Sadeleer N (2019) The precautionary principle and management of uncertainties in EU law of chemicals. Jean Monnet Paper Series - Environment and Internal market vol. 2019/1.

Delignette-Muller MD, Dutang C (2015) fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software 64(4): 1-34. URL http://www.jstatsoft.org/v64/i04/.

Dietrich E, Schulze A (2009) Statistische Verfahren zur Maschinen- und Prozessqualifikation. Carl Hanser Verlag, Munich

Ditoro DM, Zarba CS, Hansen DJ, Berry WJ, Swartz RC, Cowan CE, Pavlou SP, Allen HE, Thomas NA, Paquin PR (1991) Technical Basis for Establishing Sediment Quality Criteria for Nonionic Organic-Chemicals Using Equilibrium Partitioning. Environmental Toxicology and Chemistry 10:1541-1583.

EC - European Commission (2002) Guidance Document on Terrestrial Ecotoxicology Under Council Directive 91/414/EEC. Draft Working Document SANCO/10329/2002 rev. 2 final. Brussels, Belgium.

ECB - European Chemical Bureau (2000) Classification Criteria for the Terrestrial Environment. ECB/19/99 Add. 8 Rev. 4. Ispra, 21 July 2000.

ECHA - European Chemicals Agency (2018) IUCLID 6 - International Uniform ChemicaL Information Database. http://echa.europa.eu/

EFSA - European Food Safety Authority (2019) Open EFSA. Accessed online 2019-11-15: https://www.efsa.europa.eu/en/topics/topic/open-efsa

EFSA PPR Panel - EFSA Panel on Plant Protection Products and their Residues, Ockleford C, Adriaanse P, Berny P, Brock T, Duquesne S, Grilli S, Hernandez-Jerez AF, Bennekou SH, Klein M, Kuhl T, Laskowski R, Machera K, Pelkonen O, Pieper S, Stemmer M, Sundh I, Teodorovic I, Tiktak A, Topping CJ, Wolterink G, Craig P, de Jong F, Manachini B, Sousa P, Swarowsky K, Auteri D, Arena M, Rob S (2017) Scientific Opinion addressing the state of the science on risk assessment of plant protection products for in-soil organisms. EFSA Journal 15(2):4690-4915, 225 pp. doi:10.2903/j. efsa.2017.4690 ISSN: 1831-4732. Final version.

Environment Canada (2005) Biological test method - test for measuring emergence and growth of terrestrial plants exposed to contaminants in soil. Environmental Protection series. [Reference methods] EPS 1/RM/45 . Method Development and Application Section, Environmental Technology Centre, Environment Canada.En49-7/1-45E-PDF

European Chemicals Agency - ECHA (2017) Guidance on Information Requirements and Chemical Safety Assessment - Chapter R.7c: Endpoint specific guidance. ECHA-17-G-11-EN; 10.2823/43472; Version 3.0; June 2017

European Commission (2001) The available scientific approaches to assess the potential effects and risk of chemicals on terrestrial ecosystems. CSTEE - Scientific Committee on Toxicity, Ecotoxicity and The Environment. Opinion expressed at the 19th CSTEE Plenary Meeting. pp. 178.

European Commission (2021) EU Soil Strategy for 2030 - Reaping the benefits of healthy soils for people, food, nature and climate. COM (2021) 699 final; SWD(2021) 323 final

European Communities (2003) Technical Guidance Document on Risk Assessment. EUR 20418 EN/1. Parts I-IV. In support of Commission Directive 93/67/EEC on Risk Assessment for new notified substances, Commission Regulation (EC) No 1488/94 on Risk Assessment for existing substances, Directive 98/8/EC of the European Parliament and of the Council concerning the placing of biocidal products on the market.

Feibicke M, Vormann K, Riepert F (1999) German ad-hoc-Proposal of a Classification System for the terrestrial environment (Draft January 1999).

Frampton GK, Jänsch S, Scott-Fordsmand JJ, Römbke J, Van den Brink PJ (2006) Effects of pesticides on soil invertebrates in laboratory studies: A review and analysis using Species Sensitivity Distributions. Environmental Toxicology and Chemistry 25(9): 2480-2489

Gainer A, Bresee K, Hogan N, Siciliano SD (2019) Advancing soil ecological risk assessments for petroleum hydrocarbon contaminated soils in Canada: Persistence, organic carbon normalization and relevance of species assemblages. Science of the Total Environment 668: 400-410

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013) Bayesian Data Analysis. Third edition. Chapman and Hall/CRC, Boca Raton

González-Doncel M, Ortiz J, Izquierdo JJ, Martín B, Sánchez P, Tarazona JV (2006) Statistical evaluation of chronic toxicity data on aquatic organisms for the hazard identification: The chemicals toxicity distribution approach Chemosphere 63: 835-844

Guerra CA, Bardgett RD, Caon L, Crowther TW, Delgado-Baquerizo M, Montanarella L, Navarro LM, Orgiazzi A, Singh BK, Tedersoo L, Vargas-Rojas R, Briones MJI, Buscot F, Cameron EK, Cesarz S, Chatzinotas A, Cowan DA, Djukic I, Van den Hoogen J, Lehmann A, Maestre FT, Marín C, Reitz T, Rillig MC, Smith LC, De Vries FT, Weigelt A, Wall DH, Eisenhauer N (2021) Tracking, targeting, and conserving soil biodiversity - A monitoring and indicator system can inform policy. Science 371 (6526): 239-241

Hartel PG (2005) Microbial Processes - Environmental Factors. Encyclopedia of Soils in the Environment 448-455

Hartmann NB, Gottardo S, Sokull-Klüttgen B (2014) Review of available criteria for non-aquatic organisms within PBT/vPvB frameworks. Part II: Toxicity assessment. JRC Science and Policy Reports

Hyndman RJ, Fan Y (1996) Sample Quantiles in Statistical Packages. The American Statistician 50 (4): 361-365

ISO - International Standards Organization (1993) Soil Quality - Effects of pollutants on Earthworms (Eisenia fetida). Part 1: Determination of acute toxicity using artificial soil substrate. Geneva, Switzerland. ISO/DIS 11268-1

ISO - International Standards Organization (1996) Soil Quality - Effects of pollutants on Earthworms (Eisenia fetida, Eisenia fetida andrei). Part 2: Determination of effects on reproduction. Geneva, Switzerland. ISO/DIS 11268-2

ISO - International Standards Organization (1999) Soil Quality - Effects of Pollutants on Earthworms - Part 3: Guidance on the Determination of Effects in Field Situations. Geneva, Switzerland. ISO/DIS 11268-3

Kase R, Korkaric M, Werner I, Ågerstrand M (2016) Criteria for Reporting and Evaluating ecotoxicity Data (CRED): comparison and perception of the Klimisch and CRED methods for evaluating reliability and relevance of ecotoxicity studies. Environmental Sciences Europe 28: 7-20; DOI 10.1186/s12302-016-0073-x

Ksoll W, Schildhauer T, Beck A (2017) Open Data - Wertschöpfung im digitalen Zeitalter. Bertelsmann Stiftung, https://creativecommons.org/licenses/by-sa/4.0/.

Lanphear BP (2017) Low-level toxicity of chemicals: No acceptable levels? PLoS Biol 15(12): e2003066. https://doi.org/10.1371/journal.; pbio.2003066

Lehmann R (2012) Der Einfluss statistischer Ausreißer auf die Schätzung der natürlichen Variabilität in Daten zu Biota. Dissertation RWTH Aachen. pp. 398

Levene H (1961) Robust tests for equality of variances. In: Contributions to probability and statistics. Essays in honor of Harold Hotelling 279-292.

Millard SP (2013) EnvStats: An R Package for Environmental Statistics. Springer, New York. ISBN; 978-1-4614-8455-4.

NZEPA - New Zealand's Environmental Protection Authority (2012) Thresholds and Classifications Under the Hazardous Substances and New Organisms Act 1996. (Content as initially published March 2008)

OECD - Organisation for Economic Co-operation and Development (2000) Guidance document on aquatic toxicity testing of difficult substances and mixtures. OECD SERIES ON TESTING AND ASSESSMENT Number 23; ENV/JM/MONO (2000)6

OECD - Organisation for Economic Co-operation and Development (2000) Test No. 216: Soil Microorganisms: Nitrogen Transformation Test. OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris, https://doi.org/10.1787/9789264070226-en.

OECD - Organisation for Economic Co-operation and Development (2000) Test No. 217: Soil Microorganisms: Carbon Transformation Test. OECD Guidelines for the Testing of Chemicals, Section 2, OECD Publishing, Paris, https://doi.org/10.1787/9789264070240-en.

OECD - Organisation for Economic Co-operation and Development (2004) Earthworm Reproduction Test (Eisenia fetida / Eisenia andrei). OECD Guideline for the Testing of Chemicals 222

OECD - Organisation for Economic Co-operation and Development (2016) Test No. 220 Enchytraeid Reproduction Test.

Ökopol (Ökopol - Institute for Environmental Strategies & Beratungs- und Informationsstelle Arbeit und Gesundheit) (2004) Final report - Technical Assistance to the Commission on the implementation of the GHS. http://ec.europa.eu/enterprise/sectors/chemicals/files/ghs/ghs_impl_final_report_en.pdf

Orgiazzi A, Panagos P, Yigini Y, Dunbar MB, Gardi C, Montanarella L, Ballabio C (2016) A knowledge-based approach to estimating the magnitude and spatial patterns of potential threats to soil biodiversity. Science of the Total Environment 545-546: 11-20.

R Core Team (2020) R: A language and environment for statistical computing. Austria. URL https://www.R-project.org/. R Foundation for Statistical Computing, Vienna, Rauert C, Friesen A, Hermann G, Jöhncke U, Kehrer A, Neumann M, Prutz I, Schönfeld J, Wiemann A, Willhaus K, Wöltjen J & Duquesne S (2014) Proposal for a harmonised PBT identification across different regulatory frameworks Environmental Sciences Europe 26, 9

Renaud FG, Boxall ABA, Toy R, Robertson S (2004) Evaluation of approaches for terrestrial hazard classification. Chemosphere 57: 1697-1706

Ripley B, Lapsley M (2017) RODBC: ODBC Database Access. R package version 1.3-15. https://CRAN.R-project.org/package=RODBC

Römbke J, Gardi C, Creamer R, Miko L (2016) Soil biodiversity data: Actual and potential use in European and national legislation Applied Soil Ecology 97: 125-133

Römbke J, Jänsch S, Scheffczyk A, Craig P, Sousa JP, Martins P, Natal-da-Luz T, Scholz-Starke B (2019) UNCERTAIN - Further development of toxicity tests with soil organisms: Sustainable protection of biodiversity in soils FKZ 3718 64 402 0

Römbke J, Sousa JP (2017) Performance of laboratory tests with soil organisms - Short title: Test soils and soil organisms. Final report UBA-Project-No. 77696; ECT Project No. P16051

Saouter E, Biganzoli F, Pant R, Sala S, Versteeg D (2019) Using REACH for the EU Environmental Footprint: Building a Usable Ecotoxicity Database, Part I. Integrated Environmental Assessment and Management 15(5): 783-795

Scholz-Starke B (2013) Assessing the risks of pesticides to soil communities using terrestrial model ecosystems. PhD Thesis, RWTH Aachen University. pp. 291. http://darwin.bth.rwth-aachen.de/opus3/volltexte/2013/4825/

Scholz-Starke B (2015) Analysis of existing laboratory studies with earthworms: Influence of study design variants on ecotoxicological test results of plant protection products and veterinary pharmaceuticals. Final report Umweltbundesamt FKZ 93-104-4-17. pp. 55

Schudoma D, Michel B, Lippke A (2008) ETOX - Informationssystem Ökotoxikologie und Umweltqualitätsziele - Handbuch zur Recherche und Datenerfassung. Umweltbundesamt FG IV 2.4 Wassergefährdende Stoffe - Ökotoxikologielabor.

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (for complete samples). Biometrika 52: 591-611

Smith KEC, Rein A, Trapp S, Mayer P, Karlson UG (2002) Dynamic passive dosing for studying the biotransformation of hydrophobic organic chemicals: Microbial degradation as an example. Environmental Science & Technology 46 (4852-4860).

Stephens MA (1986) Tests based on edf statistics. In: Goodness-of-fit techniques (D'Agostino RB, Stephens MA, eds.), Marcel Dekker, New York, pp. 97-194.

Szöcs E, Muench D, Ranke J, Scott E, Stanstrup J, Allaway R (2015) Webchem-package - Chemical Information from the Web. webchem: zenodo release. Zenodo. 10.5281/zenodo.33823.

Tarazona JV, Fresno A, Aycart S, Ramos C, Veja MM, Carbonell G (2000) Assessing the potential hazard of chemical substances for the terrestrial environment. Development of hazard classification criteria and quantitative environmental indicators. Science of Total Environment 247: 151-164

Torstensson L, Pettersson I, Lundgren A (1999) Environmental hazard classification criteria for chemical substances terrestrial environment —fate in the soil and soil compartment effects. TemaNord 1999: xxx. Nordic Council of Ministers, Copenhagen

Umweltbundesamt (2019) https://webetox.uba.de/webETOX/index.do accessed 2019-10-04

UN - United Nations - Sub-Committee of Experts on the Globally Harmonized System of Classification and Labelling of Chemicals (2006) Environmental hazards, Classification criteria for the terrestrial environment

Twelfth session, 12 (p.m)-14 July 2006, Item 2 (c) of the provisional agenda. Transmitted by the expert from Spain on behalf of the group on terrestrial hazards. UN/SCEGHS/12/INF.5.

US EPA - United States Environmental Protection Agency (2012) Technical Overview of Ecological Risk Assessment Analysis Phase: Ecological Effects Characterization. Last updated on May 09, 2012

US EPA - United States Environmental Protection Agency (2019) ECOTOX User Guide: ECOTOXicology Knowledgebase System. Version 5.1. Available: http:/www.epa.gov/ecotox/ (retrieved 2018-12-17)

Van Dijk J, Gustavsson M, Dekker SC, Van Wezel AP (2021) Towards 'one substance – one assessment': An analysis of EU chemical registration and aquatic risk assessment frameworks. Journal of Environmental Management 280: 111692

Vega MM, Ramos C, Tarazona JV (1999) Statistical study on the data availability and distribution of the (eco)toxicological information suitable for the development of a hazard identification system for the terrestrial environment. In: Approaches for a hazard system for the terrestrial environment. Published by the European Chemcials Bureau, E.C.B. S.P.I. 99.80. pp 188-196

Vega MM, Vollmer G, Berggren E, Nichelatti M, Payá Pérez A (eds.) (2003) Cost and benefit analysis of the development and use of environmental hazard-effects classification criteria for terrestrial organisms. EUR Report 20689 EN.

Verbruggen EMJ, Posthumus R, Van Wezel AP (2001) Ecotoxicological serious risk concentrations for soil, sediment and (ground)water: updated proposals for first series of compounds. RIVM report 711701020. Bilthoven, the Netherlands: Nat. Inst. Public Health Environ.

Wardhaugh KG, Rodriguez-Menendez H (1988) The effects of the antiparasitic drug, ivermectin, on the development and survival of the dung-breeding fly, Orthelia cornicina (F.) and the scarabaeine dung beetles, Copris hispanus L., Bubas bubalus (Oliver) and Onitis belial. Journal of Applied Entomology 106: 381-389; https://doi.org/10.1111/j.1439-0418.1988.tb00607.x

Welch BL (1947) The generalization of Students' problem when several differnt population variances are involved. Biometrika 34(1-2): 28-35

Wheeler JR, Grist EPM, Leung KMY, Morritt D, Crane M (2002) Species sensitivity distributions: Data and model choice. Marine Pollution Bulletin 45:192-202.

Wickham H (2019) httr: Tools for Working with URLs and HTTP. R package version 1.4.1. https://CRAN.R-project.org/package=httr

A Appendices

Extensive derivative data collections within the present document that exceed the primary text frame are provided here as annexes.

A.1 Missing entries

Table annex 1: Percentages of missing entries per registration area (and combinations if a substance appears under various regulatory areas in the database) and variable.

All variable names are described in detail in chapter 3. The registration area abbreviations and their combinations stand for biocides (B), pharmaceuticals (H), plant protection products (P), REACH chemicals (R), veterinary products (V), literature sources (L), and non-specified chemicals (C).

VAR	B	B & P	B & P & R	B & P &V	B & R	C	C & P	C & R	H	H & V	L	P	P & R	P & R & V	P & V	R	R & V	V
testid	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
aicommon	46.2	0	0	0	2.9	82.5	0	58	86	0	15.7	17.3	2.2	0	0	28.2	0	35.3
aicas	3.8	0	0	0	0	21.2	0	0	0	0	0	7.6	0	0	0	7.8	0	5.8
aiiupac	46.2	0	0	0	2.9	48.1	0	36.5	0	0	0.3	17	0	0	0	0.5	0	17.7
aiec	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
testspecies	7.7	4.7	2.9	0.8	1.8	12.2	5.2	8.9	4.7	2.6	0	3.1	3.7	1.2	0.5	10.7	16.7	5.7
origin	0	e0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
testtype	30.8	4.7	20.6	0.8	51.3	16.1	22.6	60.1	17.3	23.7	1.8	5.4	6.8	2.3	4.1	31.5	0	16.1
measendpoint	50	11.6	4.9	6	8.1	6.7	4.2	39.7	3.9	2.6	0	5.5	4.6	2.5	5.5	8.4	16.7	2.8
statendpoint	3.8	7	5	0	4.5	27.6	7.3	1.7	9.8	0	0	4.5	4.6	1.2	1.4	6.4	27.8	16.2
ecotoxvalue	3.8	7	4.2	0	4.2	5.8	7.5	8	6.4	2.6	2.3	3.5	4.4	0.9	1.9	7	16.7	9.3
ecotoxmeasure	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

VAR	B	B & P	B & P & R	B & P &V	B & R	C	C & P	C & R	H	H & V	L	P	P & R	P & R & V	P & V	R	R & V	V
ecotoxunit	3.8	7	3.5	0	4.3	10	8	3.7	10.3	2.6	1	3.1	4	7.7	2.4	6.2	16.7	5.2
duration	15.4	23.3	21.1	25	5.6	11.5	5	40.5	20.7	34.2	14.1	18.6	20.6	11.1	24.1	10.7	16.7	12.5
durationunit	11.5	7	3.3	2.1	5.9	10.8	4.4	14.7	20.1	2.6	0	3.7	3.6	1.5	2.4	10.1	16.7	9.1
corg	100	100	99.6	97.9	100	98.4	99	91.1	100	100	73.4	83	89.7	90.7	86.7	89.2	100	89.8
databaserno	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
assessarea	0	0	0.4	0	0	0	0	0.6	0	0	0.2	0.4	0	0.1	0	1.2	0	0.2
testitem	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
guideline	26.9	51.2	74	98.8	86.5	21.7	52.2	27.9	16.8	26.3	87.5	69.7	73.4	88.8	68.9	64.4	44.4	30.3
operator	46.2	76.7	86.6	99.4	94.3	61.8	75.3	53.7	71.8	68.4	93.1	84.3	86.8	96.2	94.2	82.6	88.9	79.4
litsource	0	0	29.4	0	93	0	0	26.1	0	0	0	0	13.2	0	0	72.8	27.8	0
product	3.8	48.8	81.4	93.3	95.3	19.7	52.8	96.6	27.1	68.4	100	73.1	89.9	87.9	74.2	99.8	55.6	49
rowno	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
litid	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
doi	0	0	33.8	0	93	0	1.3	26.1	0	0	1.4	0.3	13.5	0.8	0	72.9	27.8	0.1
acidconst	73.1	100	100	100	100	95.9	26.4	74.7	100	0	77.3	72.1	85.4	0	100	93.8	44.4	81.5
activity	46.2	0	0	0	54.2	96.4	0	100	100	100	38.4	17.6	16.7	0	0	87.4	44.4	70.4
atmorateconst	50	0	36.9	0	50.4	85.8	81.1	66.7	86	0	23.8	23.8	9.7	0	0	75.2	44.4	59.6
boilpoint	88.5	100	100	0	97.1	72.3	83.9	58.3	100	100	75.7	71.2	23	100	100	77.7	44.4	83.7
charge	46.2	0	0	0	50.8	48.1	0	36.5	0	0	5.2	17.7	1.4	0	0	28.1	0	17.7
hdonorcount	46.2	0	0	0	50.8	48.1	0	36.5	0	0	5.2	17.7	1.4	0	0	28.1	0	17.7
hacceptcount	46.2	0	0	0	50.8	48.1	0	36.5	0	0	5.2	17.7	1.4	0	0	28.1	0	17.7

VAR	B	B & P	B & P & R	B & P &V	B & R	C	C & P	C & R	H	H & V	L	P	P & R	P & R & V	P & V	R	R & V	V
class_substance	100	100	100	100	100	99.8	100	100	100	100	95.7	97.4	100	100	100	84.1	100	100
effectspec	96.2	27.9	36.9	100	97.1	61.4	42.6	37.4	100	100	98.5	90.3	78.2	0	100	90	100	94.2
fatsol	100	100	36.9	100	97.1	98.3	100	83.6	100	100	99	98.9	90.1	100	100	99.2	100	100
henconst	46.2	0	36.9	0	100	70.8	23.7	26.7	86	0	25.7	23	9.7	0	0	83.7	44.4	59.4
logkow	46.2	0	36.9	0	93.3	49.5	0	35.3	10.1	0	25	19.1	7.5	0	0	78.6	0	19.4
meltpoint	38.5	0	36.9	0	47.5	75.2	19.5	4	74.6	0	36.1	24.4	3.5	0	0	75.9	44.4	15.7
modeaction	96.2	100	36.9	100	100	78.6	42.6	100	100	100	100	64.3	70.5	0	0	92.2	100	94.2
molweight	34.6	0	0	0	50.8	12.2	0	1.1	0	0	5.1	10.1	0	0	0	20.2	0	11.9
pubchemid	46.2	0	0	0	50.8	48.1	0	36.5	0	0	5.2	17.7	1.4	0	0	28.1	0	17.7
reldens	88.5	100	36.9	0	97.1	97.5	23.1	77.3	100	100	97.2	72.7	61.2	0	100	97.6	100	100
vaporpress	46.2	0	0	0	47.5	82.6	23.7	53.7	80.4	0	22.8	19.8	2.1	0	0	73.6	44.4	9.2
watersol	46.2	0	0	0	47.5	81.6	19.5	41.1	44.1	0	22.5	19.2	3.5	0	0	73	44.4	19.4
dissconst	96.2	100	100	100	100	78.8	100	100	58.7	100	99.5	92.1	99.6	100	100	92.2	100	67.6
kingdom	7.7	4.7	24.7	1	61.9	31.1	9	14.9	27.9	2.6	0.5	5.8	7.4	2.2	1.4	28.9	27.8	20.8
phylum	11.5	4.7	24.9	1	62	31.4	9.6	14.9	27.9	2.6	3.2	7.7	8.6	2.3	1.4	29.1	27.8	22.9
class	11.5	4.7	25.4	1	62.8	31.4	13.2	14.9	27.9	2.6	3.8	8.3	8.9	2.3	1.4	29.4	27.8	23
order	11.5	4.7	25.4	1	62.8	31.4	14.7	14.9	27.9	2.6	4.5	9.1	9.1	8.7	4.3	29.7	27.8	25
family	11.5	4.7	25.4	1	62.8	31.4	14	14.9	27.9	2.6	4.6	9	9.5	18.7	4.3	29.5	27.8	27.1
genus	11.5	4.7	26.1	1.7	62.8	31.4	18	14.9	27.9	2.6	5.8	10.3	10	28.5	5.8	29.7	27.8	28.9
species	11.5	4.7	30.4	10.8	66.5	33	21.8	16.7	29.1	2.6	12.7	16.8	16.1	36.2	22.4	33.7	27.8	31.8
subspecies	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

VAR	B	B & P	B & P & R	B & P &V	B & R	C	C & P	C & R	H	H & V	L	P	P & R	P & R & V	P & V	R	R & V	V
variety	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
effecttype	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
relevance_uba	73.1	34.9	35.3	10	64.1	28.5	43.4	67.5	26.3	57.9	14.5	22.9	19.5	16.6	25.1	47.5	27.8	25.8
standardtest_uba	76.9	34.9	37	10.2	65.1	29.2	44.4	67.5	29.9	57.9	19.6	26	22.4	17.5	25.3	49.8	44.4	28.1
level_orga	50	11.6	6.1	6	8.1	8	6.9	39.7	3.9	2.6	0.1	6	4.6	3.1	5.5	8.5	16.7	2.8
subgroup	26.9	51.2	85	98.3	68.3	48.4	78.4	39.9	61.2	76.3	92.9	86.7	87.3	78.1	70.4	72.2	77.8	72.2

A.2 Candidate lists – 30 upper most toxic substances 30

Table annex 2: Candidate list of criterion group 1 "plants acute", upper 30 most toxic candidates

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
145701-23-1	9	≤ 0,0001	Beta vulgaris	florasulam	verytoxic01	verytoxic01	verytoxic01	PESTI				h400 & h410	h400 & h410
128621-72-7	6	≤ 0,0001	Abutilon theophrasti	carfentrazone	verytoxic02	verytoxic02	verytoxic02	CROSS					
128639-02-1	2	≤ 0,0001	Abutilon theophrasti	carfentrazone-ethyl	verytoxic03	verytoxic03	verytoxic03	PESTI				h400 & h410	h400 & h410
129630-19-9	3	≤ 0,0001	Beta vulgaris	pyraflufen-ethyl	verytoxic04	verytoxic04	verytoxic04	PESTI				h400 & h410	h400 & h410
74223-64-6	21	≤ 0,0001	Inula helenium	metsulfuron-methyl	verytoxic05	verytoxic05	verytoxic05	PESTI				h400 & h410	h400 & h410
6753-47-5	1	≤ 0,0001	Glycine max	picloram-tripromine	verytoxic06	verytoxic06	verytoxic06	LITRA					h400 & h410 & h412

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
144550-36-7	1	≤ 0,0001	Sinapis alba	iodosulfuron-methyl-sodium	verytoxic07	verytoxic07	verytoxic07	PESTI				h400 & h410	h400 & h410
101200-48-0	17	≤ 0,0001	Cucumis sativa	tribenuron-methyl	verytoxic08	verytoxic08	verytoxic08	PESTI				h400 & h410	h400 & h410
120923-37-7	3	≤ 0,0001	Amaranthus retroflexus	amidosulfuron	verytoxic09	verytoxic09	verytoxic09	PESTI				h400 & h410	h400 & h410
1071-83-6	18	≤ 0,001	Helianthus annuus	glyphosphate	verytoxic10	verytoxic10	verytoxic10	PESTI				h411 & h413	h411 & h413
141776-32-1	3	≤ 0,001	Camelina sativa	sulfosulfuron	verytoxic11	verytoxic11	verytoxic11	PESTI				h400 & h410	h400 & h410
99129-21-2	30	≤ 0,001	Setaria faberi	clethodim	verytoxic12	verytoxic12	verytoxic12	PESTI				h412	h412
100646-51-3	4	≤ 0,001	Echinochloa crus-galli	quizalofop-p-ethyl	verytoxic13	verytoxic13	verytoxic13	PESTI					h400 & h410 & h413
145026-88-6	3	≤ 0,001	Camelina sativa	flucarbazone	verytoxic14	verytoxic14	verytoxic14	LITRA					
122931-48-0	12	≤ 0,001	Brassica napus	rimsulfuron	verytoxic15	verytoxic15	verytoxic15	PESTI					h400 & h410
106040-48-6	7	≤ 0,001	Myosoton aquaticum	tribenuron	verytoxic16	verytoxic16	verytoxic16	PESTI					h410
76578-14-8	7	≤ 0,001	Echinochloa crus-galli	quizalofop-ethyl	verytoxic17	verytoxic17	verytoxic17	LITRA					h400
141112-29-0	30	≤ 0,001	Brassica oleracea	isoxaflutole	verytoxic18	verytoxic18	verytoxic18	PESTI				h400 & h410	h400 & h410
181274-17-9	31	≤ 0,001	Setaria viridis	flucarbazone-sodium	verytoxic19	verytoxic19	verytoxic19	LITRA					
122548-33-8	4	≤ 0,001	Daucus carota	imazosulfuron	verytoxic20	verytoxic20	verytoxic20	PESTI					h400 & h410

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
123343-16-8	21	≤ 0,001	Zea mays	pyrithiobac-sodium salt	verytoxic21	verytoxic21	verytoxic21	LITRA					h400 & h410 & h413
81335-77-5	11	≤ 0,001	Camelina sativa	imazethapyr	verytoxic22	verytoxic22	verytoxic22	LITRA					h400 & h410
943832-60-8	1	≤ 0,001	Daucus carota	halauxifen	verytoxic23	verytoxic23	verytoxic23	CROSS					
943831-98-9	2	≤ 0,001	Glycine max	halauxifen-methyl	verytoxic24	verytoxic24	verytoxic24	PESTI					h400 & h410 & h411
15545-48-9	5	≤ 0,001	Lactuca sativa	chlortoluron	verytoxic25	toxic01	verytoxic25	PESTI				h400 & h410	h400 & h410
111991-09-4	17	≤ 0,001	Oryza sativa	nicosulfuron	verytoxic26	toxic02	verytoxic26	PESTI					h400 & h410 & h411 & h412
123342-93-8	24	≤ 0,001	Amaranthus hybridus	pyrithiobac	verytoxic27	toxic03	verytoxic27	LITRA					h400 & h410
79277-27-3	1	≤ 0,001	Beta vulgaris	harmony	verytoxic28	toxic04	verytoxic28	PESTI				h400 & h410	h400 & h410
87392-12-9	26	≤ 0,001	Cucumis sativa	s-metolachlor	verytoxic29	toxic05	verytoxic29	PESTI				h400 & h410	h400 & h410
94125-34-5	4	≤ 0,001	Lactuca sativa	prosulfuron	verytoxic30	toxic06	verytoxic30	PESTI				h400 & h410	h400 & h410

CAS = CAS number; N = n data; VAL = ecotox value rounded [mg ai/kg sdw]; SENS = most sensitive species; NAME = common name; RGEO = rank geomean approach; RQUA = rank quantile approach; RSIG=rank sigtest approach; REG = registration area; PBT = PBT status; SVHC = SVHC status; POP = POP status; CLHAR = ECHA CL status harmonized; CLSEL = ECHA CL status self-classified

Table annex 3: Candidate list of criterion group 2 “plants chronic”, upper 30 most toxic candidates

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
144550-36-7	4	≤ 0,00001	Lactuca sativa	iodosulfuron-methyl-sodium	verytoxic01	verytoxic01	verytoxic01	PESTI				h400 & h410	h400 & h410
64902-72-3	20	≤ 0,00001	Allium cepa	chlorsulfuron	verytoxic02	verytoxic02	verytoxic02	PESTI				h400 & h410	h400 & h410
122548-33-8	27	≤ 0,00001	Beta vulgaris	imazosulfuron	verytoxic03	verytoxic03	verytoxic03	PESTI					h400 & h410
79277-27-3	14	≤ 0,00001	Lycopersicon esculentum	harmony	verytoxic04	verytoxic04	verytoxic04	PESTI				h400 & h410	h400 & h410
81334-34-1	14	≤ 0,00001	Cucumis sativa	imazapyr	verytoxic05	verytoxic05	verytoxic05	LITRA				h400 & h410	h400 & h410 & h412
101200-48-0	51	≤ 0,0001	Beta vulgaris	tribenuron-methyl	verytoxic06	verytoxic06	verytoxic06	PESTI				h400 & h410	h400 & h410
128621-72-7	2	≤ 0,0001	Allium cepa	carfentrazone	verytoxic07	verytoxic07	verytoxic07	CROSS					
173159-57-4	27	≤ 0,0001	Raphanus sativus	foramsulfuron	verytoxic08	verytoxic08	verytoxic08	PESTI					h412
422556-08-9	19	≤ 0,0001	Daucus carota	pyroxsulam	verytoxic09	verytoxic09	verytoxic09	PESTI				h400 & h410	h400 & h410
141112-29-0	30	≤ 0,0001	Lactuca sativa	isoxaflutole	verytoxic10	verytoxic10	verytoxic10	PESTI				h400 & h410	h400 & h410
100784-20-1	11	≤ 0,0001	Lactuca sativa	halosulfuron-methyl	verytoxic11	verytoxic11	verytoxic11	LITRA				h400 & h410	h400 & h410
145701-21-9	11	≤ 0,0001	Gossypium hirsutum	diclosulam	verytoxic12	verytoxic12	verytoxic12	LITRA					h400 & h410

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
147150-35-4	10	≤ 0,0001	Raphanus sativus	cloransulam	verytoxic13	verytoxic13	verytoxic13	LITRA					h400 & h410
104040-78-0	19	≤ 0,0001	Raphanus sativus	flazasulfuron	verytoxic14	verytoxic14	verytoxic14	PESTI				h400 & h410	h400 & h410
950782-86-2	30	≤ 0,0001	Beta vulgaris	indaziflam	verytoxic15	verytoxic15	verytoxic15	LITRA					h400 & h410
122931-48-0	7	≤ 0,0001	Brassica napus	rimsulfuron	verytoxic16	verytoxic16	verytoxic16	PESTI					h400 & h410
74222-97-2	5	≤ 0,0001	Beta vulgaris	sulfometuron methyl	verytoxic17	verytoxic17	verytoxic17	LITRA					h400 & h410
141776-32-1	4	≤ 0,0001	Raphanus sativus	sulfosulfuron	verytoxic18	verytoxic18	verytoxic18	PESTI				h400 & h410	h400 & h410
372137-35-4	32	≤ 0,0001	Glycine max	saflufenacil	verytoxic19	verytoxic19	verytoxic19	LITRA					h400 & h410
134605-64-4	19	≤ 0,0001	Raphanus sativus	butafenacil	verytoxic20	verytoxic20	verytoxic20	REACH					h400 & h410
74223-64-6	1	≤ 0,0001	Fallopia convolvulus	metsulfuron-methyl	verytoxic21	verytoxic21	verytoxic21	PESTI				h400 & h410	h400 & h410
188489-07-8	16	≤ 0,0001	Lactuca sativa	flufenpyr-ethyl	verytoxic22	verytoxic22	verytoxic22	LITRA					
103361-09-7	23	≤ 0,0001	Cucumis sativa	flumioxazin	verytoxic23	verytoxic23	verytoxic23	PESTI				h400 & h410	h400 & h410
115136-53-3	22	≤ 0,0001	Brassica oleracea	imazapic-ammonium	verytoxic24	verytoxic24	verytoxic24	LITRA					
1918-16-7	15	≤ 0,0001	Raphanus sativus	propachlor	verytoxic25	verytoxic25	verytoxic25	PESTI				h400 & h410	h400 & h410

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
317815-83-1	51	≤ 0,0001	Brassica napus	thiencarbazone-methyl	verytoxic26	verytoxic26	verytoxic26	PESTI				h400 & h410	h400 & h410
21293-29-8	1	≤ 0,0001	Phaseolus vulgaris	2,4-pentadienoic acid, 5-(1-hydroxy-2,6,6-trimet	verytoxic27	verytoxic27	verytoxic27	LITRA					
858954-83-3	17	≤ 0,001	Glycine max	aminocyclopyrachlor-methyl	verytoxic28	verytoxic28	verytoxic28	LITRA					
117337-19-6	14	≤ 0,001	Cucumis sativa	fluthiacet-methyl	verytoxic29	verytoxic29	verytoxic29	LITRA					h400
104206-82-8	43	≤ 0,001	Lactuca sativa	mesotrione	verytoxic30	verytoxic30	verytoxic30	PESTI				h400 & h410	h400 & h410

CAS = CAS number; N = n data; VAL = ecotox value rounded [mg ai/kg sdw]; SENS = most sensitive species; NAME = common name; RGEO = rank geomean approach; RQUA = rank quantile approach; RSIG=rank sigtest approach; REG = registration area; PBT = PBT status; SVHC = SVHC status; POP = POP status; CLHAR = ECHA CL status harmonized; CLSEL = ECHA CL status self-classified

Table annex 4: Candidate list of criterion group 3 “soil macro acute”, upper 30 most toxic candidates

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
91465-08-6	17	≤ 0,0001	Phytoseiulus persimilis	lambda-cyhalothrin	verytoxic01	verytoxic01	verytoxic01	PESTI&REACH				h400 & h410	h400 & h410
52315-07-8	13	≤ 0,001	Eisenia fetida	cypermethrin	verytoxic02	verytoxic02	verytoxic02	PESTI				h400 & h410	h400 & h410
135410-20-7	5	≤ 0,01	Eisenia fetida	acetamiprid	verytoxic03	verytoxic03	verytoxic03	PESTI				h400	h400 & h412

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
115-29-7	14	≤ 0,01	Eisenia fetida	endosulfan	verytoxic04	verytoxic04	verytoxic04	PESTI			POP	h400 & h410	h400 & h410
116-06-3	12	≤ 0,01	Lumbricus rubellus	aldicarb	verytoxic05	verytoxic05	verytoxic05	PESTI				h400 & h410	h400 & h410
61546-00-7	1	≤ 0,01	Eisenia fetida		verytoxic06	verytoxic06	verytoxic06	LITRA					
138261-41-3	13	≤ 0,01	Eisenia fetida	imidacloprid	verytoxic07	verytoxic07	verytoxic07	PESTI&REACH&VETME				h400 & h410	h400 & h410
12789-03-6	6	≤ 0,01	Eisenia fetida	chlordane	verytoxic08	verytoxic08	verytoxic08	LITRA					h400 & h410
944-22-9	2	≤ 0,01	Lumbricus rubellus	fonofos	verytoxic09	verytoxic09	verytoxic09	PESTI				h400 & h410	h400 & h410
87-86-5	57	≤ 0,1	Eisenia fetida	pentachlorphenol	verytoxic10	verytoxic10	verytoxic10	PESTI				h400 & h410	h400 & h410 & h413
52918-63-5	11	≤ 0,1	Lumbricus rubellus	deltamethrin	verytoxic11	verytoxic11	verytoxic11	PESTI&VETME				h400 & h410	h400 & h410
150824-47-8	3	≤ 0,1	Eisenia fetida	nitenpyram	verytoxic12	verytoxic12	verytoxic12	LITRA					
121-75-5	5	≤ 0,1	Lumbricus rubellus	malathion	verytoxic13	verytoxic13	verytoxic13	PESTI				h400 & h410	h400 & h410
210880-92-5	12	≤ 0,1	Eisenia fetida	clothianidin	verytoxic14	verytoxic14	verytoxic14	CROSS&PESTI				h400 & h410	h400 & h410
63-25-2	60	≤ 0,1	Lumbricus rubellus	carbaryl	verytoxic15	verytoxic15	verytoxic15	PESTI				h400 & h410	h400 & h410
1563-66-2	13	≤ 0,1	Eisenia fetida	carbofuran	verytoxic16	verytoxic16	verytoxic16	PESTI				h400 & h410	h400 & h410

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
188589-32-4	1	≤ 0,1	Eisenia fetida		verytoxic17	verytoxic17	verytoxic17	LITRA					
141318-03-8	1	≤ 0,1	Eisenia fetida		verytoxic18	verytoxic18	verytoxic18	LITRA					
29232-93-7	10	≤ 0,1	Lumbricus rubellus	pirimiphos-methyl	verytoxic19	verytoxic19	verytoxic19	PESTI				h400 & h410	h400 & h410
111988-49-9	5	≤ 0,1	Eisenia fetida	thiacloprid	verytoxic20	verytoxic20	verytoxic20	PESTI				h400 & h410	h400 & h410
143-33-9	1	≤ 0,1	Eisenia fetida	sodium cyanide	verytoxic21	verytoxic21	verytoxic21	REACH					h400 & h410
151-50-8	1	≤ 0,1	Eisenia fetida	potassium cyanide	verytoxic22	verytoxic22	verytoxic22	REACH					h400 & h410
51-28-5	1	≤ 0,1	Eisenia fetida	2,4-dinitrophenol	verytoxic23	verytoxic23	verytoxic23	REACH				h400 & h410	h400 & h410
100-02-7	11	≤ 0,1	Eisenia fetida	4-nitrophenol	verytoxic24	verytoxic24	verytoxic24	REACH					
114-26-1	14	≤ 0,1	Lumbricus rubellus	propoxur	verytoxic25	verytoxic25	verytoxic25	PESTI				h400 & h410	h400 & h410
51218-49-6	4	≤ 0,1	Xenylla welchi	pretilchlor	verytoxic26	verytoxic26	verytoxic26	LITRA					h400 & h410
950-37-8	5	≤ 1	Eisenia fetida	methidathion	verytoxic27	verytoxic27	verytoxic27	PESTI				h400 & h410	h400 & h410
56-38-2	3	≤ 1	Lumbricus rubellus	parathion	verytoxic28	verytoxic28	verytoxic28	PESTI				h400 & h410	h400 & h410
57-74-9	5	≤ 1	Eisenia fetida	chlordan	verytoxic29	verytoxic29	verytoxic29	PESTI			POP	h400 & h410	h400 & h410

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
10605-21-7	57	≤ 1	Eisenia fetida	derosal	verytoxic30	verytoxic30	verytoxic30	PESTI				h400 & h410	h400 & h410

CAS = CAS number; N = n data; VAL = ecotox value rounded [mg ai/kg sdw]; SENS = most sensitive species; NAME = common name; RGEO = rank geomean approach; RQUA = rank quantile approach; RSIG=rank sigtest approach; REG = registration area; PBT = PBT status; SVHC = SVHC status; POP = POP status; CLHAR = ECHA CL status harmonized; CLSEL = ECHA CL status self-classified

Table annex 5: Candidate list of criterion group 5 “soil micro chronic”, upper 30 most toxic candidates

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
857036-77-2	1	≤ 0,001	mixed population		verytoxic01	verytoxic01	verytoxic01	HUMME					h400 & h410
10592-13-9	11	≤ 0,001	Glomus intraradices	doxycycline hydrochloride	verytoxic02	verytoxic02	verytoxic02	LITRA					h400 & h410
443913-73-3	1	≤ 0,01	mixed population		verytoxic03	verytoxic03	verytoxic03	HUMME					h410
81335-77-5	7	≤ 0,01	Glomus intraradices	imazethapyr	verytoxic04	verytoxic04	verytoxic04	LITRA					h400 & h410
123997-26-2	1	≤ 0,1	mixed population	eprinomectin	verytoxic05	verytoxic05	verytoxic05	VETME					h400 & h410
81335-37-7	1	≤ 0,1	Glomus intraradices	imazaquin	verytoxic06	verytoxic06	verytoxic06	LITRA					h400 & h410 & h412
298-46-4	3	≤ 0,1	Glomus intraradices	carbamazepine	verytoxic07	verytoxic07	verytoxic07	LITRA					h410 & h412
145701-21-9	1	≤ 0,1	Sclerotinia minor	diclosulam	verytoxic08	verytoxic08	verytoxic08	LITRA					h400 & h410

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
40487-42-1	8	≤ 0,1	Glomus intraradices	pendimethalin	verytoxic09	toxic01	verytoxic09	PESTI				h400 & h410	h400 & h410
1264-72-8	1	≤ 0,1	mixed population		verytoxic10	toxic02	verytoxic10	VETME					
103361-09-7	1	≤ 1	Sclerotinia minor	flumioxazin	verytoxic11	harmful01	verytoxic11	PESTI				h400 & h410	h400 & h410
60207-90-1	4	≤ 1	Glomus claroideum	propiconazole	toxic01		verytoxic12	PESTI				h400 & h410	h400 & h410
1702-17-6	1	≤ 1	Sclerotinia sclerotiorum	3,6-dichloropicolinic acid	toxic02		verytoxic13	PESTI				h410	h410 & h411
107534-96-3	3	≤ 1	Sclerotinia minor	tebuconazole	toxic03		verytoxic14	BIOCI&PESTI&REACH				h400 & h410	h400 & h410 & h411
175013-18-0	1	≤ 1	Sclerotinia minor	pyraclostrobin	toxic04		toxic01	PESTI					h400 & h410
10605-21-7	6	≤ 1	Glomus claroideum	derosal	toxic05		toxic02	PESTI				h400 & h410	h400 & h410
114369-43-6	1	≤ 1	Thanatephorus cucumeris	fenbuconazole	toxic06		toxic03	PESTI				h400 & h410	h400 & h410
131361-18-7	1	≤ 1	Thanatephorus cucumeris		toxic07		toxic04	LITRA					
148-79-8	1	≤ 1	Thanatephorus cucumeris	thiabendazole	toxic08		toxic05	PESTI				h400 & h410	h400 & h410

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
1689-84-5	7	≤ 1	Sclerotinia sclerotiorum	bromoxynil	toxic09		toxic06	PESTI				h400 & h410	h400 & h410
17804-35-2	3	≤ 1	Thanatephorus cucumeris	benlate 50	toxic10		toxic07	PESTI				h400 & h410	h400 & h410
1897-45-6	4	≤ 1	Thanatephorus cucumeris	chlorothalonil	toxic11		toxic08	PESTI				h400 & h410	h400 & h410
23564-05-8	1	≤ 1	Thanatephorus cucumeris	thiophanate-methyl	toxic12		toxic09	PESTI				h400 & h410	h400 & h410
36734-19-7	1	≤ 1	Thanatephorus cucumeris	rovral (iprodione)	toxic13		toxic10	PESTI				h400 & h410	h400 & h410
39148-24-8	1	≤ 1	Thanatephorus cucumeris	fosetyl-aluminium	toxic14		toxic11	PESTI					
5234-68-4	1	≤ 1	Thanatephorus cucumeris	carboxin	toxic15		toxic12	PESTI				h400 & h410	h400 & h410
57018-04-9	1	≤ 1	Thanatephorus cucumeris	tolclofos-methyl	toxic16		toxic13	PESTI				h400 & h410	h400 & h410
63284-71-9	1	≤ 1	Thanatephorus cucumeris	5-pyrimidinemethanol, .alpha.-(2-chlorophenyl)-.	toxic17		toxic14	LITRA					

CAS	N	VAL	SENS	NAME	RGEO	RQUA	RSIG	REG	PBT	SVHC	POP	CLHAR	CLSEL
82-68-8	2	≤ 1	Thanatephorus cucumeris	pentachloronitrobenzene	toxic18		toxic15	PESTI				h400 & h410	h400 & h410
94361-06-5	1	≤ 1	Thanatephorus cucumeris	cyproconazole	toxic19		toxic16	PESTI				h400 & h410	h400 & h410

CAS = CAS number; N = n data; VAL = ecotox value rounded [mg ai/kg sdw]; SENS = most sensitive species; NAME = common name; RGEO = rank geomean approach; RQUA = rank quantile approach; RSIG=rank sigtest approach; REG = registration area; PBT = PBT status; SVHC = SVHC status; POP = POP status; CLHAR = ECHA CL status harmonized; CLSEL = ECHA CL status self-classified

B Supplementary material (external documents)

The descriptive and explorative analyses of the database yielded a vast amount of data tables and figures. This supplementary material has been made available to the UBA as additional documents accompanying this project report. These materials consist of the complete candidate lists of all criterion groups ("candidate lists"), the ecotoxicity distributions of the assessment areas ("ecotox_distributions_assessment_areas"), the ecotoxicity distributions of all criterion groups ("ecotox_distributions_criterion_groups"), frequency tables/pie charts/word clouds of the complete data and all groups ("frequencies"), the harmonization tables ("harmonization"), as well as the species sensitivity distributions of all groups ("ssd")(see Table annex 6). In addition, further information, such as databases, a register of raw data tables, R-scripts, and complete workshop documentation, which were used during the project, have been made available to UBA for further internal use and are in most cases confidential.

Table annex 6: Short structure of supplementary materials made available as additional external documents accompanying this project report

main folder	subfolder	short description of content
candidate_lists		candidate list of all groups with harmonized plus notified ECHA C&L entries
candidate_lists	only harmonized	candidate list of all groups with only harmonized ECHA C&L entries
ecotox_distributions_assessment_areas		multiple documents with ecotoxicity distributions of all assessment areas
ecotox_distributions_criterion_groups		multiple documents with ecotoxicity distributions of all criterion groups
frequencies	complete_data	multiple documents with frequencies of the complete data for all CAS numbers in prosoildat
frequencies	group1	multiple documents with frequencies of criterion group 1
frequencies	group2	multiple documents with frequencies of criterion group 2
frequencies	group3	multiple documents with frequencies of criterion group 3
frequencies	group4	multiple documents with frequencies of criterion group 4
frequencies	group5	multiple documents with frequencies of criterion group 5
harmonization		all harmonization tables applied in this project
ssd	group1	multiple documents with species sensitivity distributions of criterion group 1

main folder	subfolder	short description of content
ssd	group2	multiple documents with species sensitivity distributions of criterion group 2
ssd	group3	multiple documents with species sensitivity distributions of criterion group 3
ssd	group4	multiple documents with species sensitivity distributions of criterion group 4
ssd	group5	multiple documents with species sensitivity distributions of criterion group 5