European Commission

JRC TECHNICAL REPORTS

Environmental Footprint: Update of Life Cycle Impact Assessment Methods – Ecotoxicity freshwater, human toxicity cancer, and non-cancer

Saouter E, Biganzoli F, Ceriani L, Versteeg D, Crenna E, Zampori L, Sala S, Pant R

Joint Research Centre

This publication is a Technical report by the Joint Research Centre (EC-JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact information
Name: Serenella Sala
Address: European Commission – EC-JRC, via Fermi 2749, 21027 Ispra, Italy
Email: serenella.sala@ec.europa.eu
Tel.: +39 0332 786417

EU Science Hub
https://ec.europa.eu/EC-JRC

How to cite this report: Saouter, E., Biganzoli, F., Ceriani, L., Versteeg, D., Crenna, E., Zampori, L., Sala, S, Pant, R. (2020). Environmental Footprint: Update of Life Cycle Impact Assessment Methods – Ecotoxicity freshwater, human toxicity cancer, and non-cancer. EUR 29495 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-17143-0, doi:10.2760/300987, JRC114227.

This report is the re-edition of the JRC Technical Report: Saouter, E., Biganzoli, F., Ceriani, L., Versteeg, D., Crenna, E., Zampori, L., Sala, S, Pant, R. (2018). Environmental Footprint: Update of Life Cycle Impact Assessment Methods – Ecotoxicity freshwater, human toxicity cancer, and non-cancer. EUR 29495 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-79-98182-1, doi: 10.2760/178544, EC-JRC114227.

Contents

Acknowledgements

Authors

Saouter Erwan

Biganzoli Fabrizio

Ceriani Lidia

Versteeg Donald

Crenna Eleonora

Zampori Luca

Sala Serenella

Pant Rana

Abstract

In 2011 the European Commission Joint Research Centre (EC-JRC) published the International Reference Life Cycle Data System (ILCD) Handbook recommendations on the use of Impact Assessment models for use in LCA (EC-JRC, 2011a). This created the basis for the Product and Organisation Environmental Footprint (PEF/OEF) recommendations for impact categories and models as per Recommendation 2013/179/EU on the use of common methods to measure and communicate the life cycle environmental performance of products and organisations (EC, 2013a). This Commission Recommendation is expected to contribute to the Building the Single Market for Green Products (EC, 2013b) by supporting a level playing field regarding the measurement of environmental performance of products and organisations.

During the PEF pilot phase (from 2013 to 2018), the model retained and recommended for assessing the impact of elementary flows on freshwater ecotoxicity and human cancer and non-cancer toxicity was the model USEtox® 1.01. However, due to the difficulties encountered in using the model and in interpretation the results, the PEF Technical Advisory Board (TAB) has decided not to include these three impact categories in the list of mandatory impact categories to be used for hotspot analysis and for communication to consumer or to business.

The EU Commission Joint Research Centre was then mandated by DG Environment to conduct an in-depth evaluation of the model and data used to calculate characterisation factors (CFs) and to come with a proposal to 1) address the issue reported by the Pilots and 2) increase the number of available characterisation factors.

Using the physicochemical and toxicity data available in the REACH, EFSA and PPDB database, and building on the feedback collected during a PEF stakeholder workshop organised in February 2018 and on the preliminary outcomes of the UNEP-SETAC Pellston workshop organised in June 2018 (UNEP- SETAC, 2018), EC-JRC has calculated new freshwater ecotoxicity characterisation factors for 6011 substances, 3423 CFS for human toxicity non-cancer and 621 CFs for human toxicity cancer.

The freshwater ecotoxicity, human cancer and non-cancer impact categories are recommended to be used in EF context, level of recommendation III.

The report describes the methodology followed to generate those new characterisation factors. Furthermore, a contribution analysis has been performed comparing the contribution to this new CFs versus old ones used in the PEF pilots.

1 Introduction

One of the goals of a life cycle assessment (LCA) is to estimate the potential impacts on ecosystems and human health of the manufacturing, use and disposal of products or services due the consumption of natural resources and the emission of substances into air, soil and aquatic environments (ISO, 2006a). LCA methodology, developed in the late 60's, was focusing mainly on energy flow, use of non-renewable energy and greenhouse gas emissions (GHG). Steadily, new impact categories have been added: i.e. ozone depletion, acidification, eutrophication, resource depletion, freshwater ecotoxicity and human toxicity, land use, etc. (EC-JRC, 2011b, 2011a). Each impact category is relying on models, which link the emissions or resource used (inventory phase) to an impact along a cause-effect chain (impact assessment phase). For the calculation of each impact category, different models are available. A review of the main LCA impact categories can be found in (EC-JRC, 2011b).

Over the years, several models for toxicity-related impact categories have been developed by different research groups: e.g. Caltox (McKone & Enoch, 2002), USES-LCA (Huijbregts et al., 2001; Van Zelm et al., 2009), TRACI (Bare, 2011), IMPACT 2002 + (Jolliet et al., 2003), EDIP 2003 (Hauschild & Potting, 2005), MEEuP (Kemna et al., 2005). These models, based on different assumptions and algorithms lead to different results – up to few orders of magnitude (Hauschild et al., 2008) - preventing a direct comparison between studies. In order to overcome intrinsic differences of the models and capitalising on the available knowledge, a consensus model – USEtox®- has been developed in the context of the UNEP-SETAC (United Nations Environmental Programme – Society of Environmental Toxicology and Chemistry) Life Cycle Initiative (Hauschild et al., 2008; Rosenbaum et al., 2008). USEtox® aims at assessing the potential impact of substances on aquatic freshwater and human using a multimedia fate modelling to estimate substance distribution in various environmental compartments. USEtox®, in its version 1.01, has been included in the ILCD recommendations (EC-JRC, 2011a), and consequently in the context of the EU Commission Product & Organization Environmental Footprint (PEF/OEF) (EC, 2013a). So far, this model has been considered by the LCA community as the most consensual for comparing the potential impact of substance emissions on human health and aquatic ecosystems (Henderson et al., 2011). This is a tier 1 model that helps identify the 10 - 20 most contributing substances in a life cycle inventory (Rosenbaum, 2015). Once this is done, further data gathering may be needed to confirm the initial outcome of the model.

In 2011, the EC-JRC published the International Reference Life Cycle Data System (ILCD) Handbook recommendations on the use of Impact Assessment models to be used in LCA (EC-JRC, 2011a). This created the basis for the Product and Organisation Environmental Footprint (PEF/OEF) recommendations for impact categories and models as per Recommendation 2013/179/EU on the use of common methods to measure and communicate the life cycle environmental performance of products and organisations (EC, 2013a). This Commission Recommendation is expected to contribute to the Building the Single Market for Green Products (EC, 2013b) by supporting a level playing field regarding the measurement of environmental performance of products and organisations.

In the context of the EF (Environmental Footprint), the model retained and recommended for assessing the impact of elementary flows on freshwater aquatic ecosystems and human cancer and non-cancer toxicity was the model USEtox® 1.01.

Since its release, USEtox® 1.01 has been widely used by research organizations and consulting firms. However, in the context of the EF, the model has been systematically used and evaluated by several sectors of industries (25 pilots) for the purpose of comparison of toxicity impacts between products.

In January 2015, the European Commission has organized a workshop with all the EF pilots that have used the USEtox® 1.01 model in their screening studies. The main conclusions from this meeting were:

- The experience of using the USEtox® 1.01 model by all pilot members has revealed some limitations. The model was not criticized as such, but rather the outcomes of the calculation (i.e. when inventory are to be multiplied by characterisation factors (CFs)).

- The model has been considered lacking of transparency and complex when there is the need of calculating new characterisation factors.

- The input data (physicochemical, half-life and toxicity) that have been used to calculate the CFs provided with the model (about 3000) were source of significant controversies and criticisms. For many substances, fate and effect data, as currently used in USEtox®, were not considered suitable.

- Most EF Pilots recommended not to use the model further before agreement is reached on the selection of input data.

- More alignment should be found between input data used to run USEtox® and data used for risk assessment purpose. In many cases industry noticed differences in e.g. physicochemical and toxicity data reported in REACH dossiers and used in USEtox®.

- For metals, the result of the UNEP-SETAC and Metal industry workshop (Diamond et al., 2010) must be implemented before CFs are calculated for metal compounds. Furthermore, metal essentiality was pointed out as an important modelling issue/gap to be addressed.

In December 2016, the PEF Technical Advisory Board (TAB) has decided not to include the freshwater ecotoxicity, human cancer and human non-cancer toxicity impact categories in the list of impact categories to be communicated or used for the identification of most relevant impact categories, life cycle stages and processes.

After the EF USEtox® workshop in January 2015, the EC-JRC was mandated by DG environment to conduct an in-depth evaluation of the USEtox® model, including the data used to calculate CFs, striving towards providing a proposal to: 1) address the issue reported by the Pilots, and 2) increase the number of available CFs.

EC-JRC has conducted the investigation gathering all the internal expertise available at EC-JRC on exposure modelling and toxicity assessment of substances. The EC-JRC unit in charge of LCA / EF has therefore worked closely with the Institute of Health and Consumer Protection, and particularly the toxicological unit and the exposure modelling unit. Additional support has been obtained by involving in the study international recognized experts in the fields of substance fate modelling and risk assessment.

On 14th February 2018, the EU Commission has organised an EF and stakeholder workshop to collect feedback and suggestions toward an agreement on the data selection procedure. A draft technical report containing the detailed background work performed by EC-JRC was shared ahead of the meeting with stakeholders involved in the EU Commission Environmental Footprint (EF) pilot activities.

This initial investigation has led to the publication of three peer-reviewed articles:

— Improving substance information in USEtox®, part 1: Discussion on data and approaches for estimating freshwater ecotoxicity effect factors (Saouter et al., 2017a).

— Improving substance information in USEtox®, part 2: Data for estimating fate and ecosystem exposure factors (Saouter et al., 2017b)

— Estimating Substance Ecotoxicity in EU Ecolabel and in EU Product Environmental Footprint (Saouter et al., 2018)

On June 24th, the 'Global Guidance for Life Cycle Impact Assessment Indicators and Methods' (Frischknecht & Jolliet, 2016) brought together approximately 40 experts (including EC-JRC) from all over the world to a 5-day workshop (the Pellston workshop (UNEP - SETAC, 2018)) to address environmental life cycle impact assessment indicators covering among others the following topics:

— Eco-toxicity;

— Human toxicity (including indoor);

The workshop was co-organized by the Life Cycle Initiative (hosted by UN Environment) in collaboration with the Society of Environmental Toxicology and Chemistry (SETAC) aims at providing scientific grounds for harmonized environmental impact indicators, which are suited for use in life cycle assessment studies. The EU Environmental Footprint has built its list of impact assessment categories on the previous work and recommendations of the Life Cycle Initiative and wants that new recommendations from the 2018 Pellston workshop can be reflected in futurs activities, as appropriate.

The key expected outcome of the follow up of the workshop is a set of characterization factors (CFs) representing acidification and eutrophication, freshwater ecotoxicity, human toxicity, natural resources (mineral primary resources), and ecosystem services with focus on soil quality. The news recommendations for the substance toxicity impact categories (both human and aquatic toxicity) require some interventions on the USEtox® model and on how some input factors are derived. Those modifications are expected to be implement in the course of 2019 and new Characterization factors will released then by the USEtox® team.

Although the final report from 'Global Guidance for Life Cycle Impact Assessment Indicators and Methods' is to be expected mid-2019, the EU Commission has decided to implement as much as possible the workshop recommendations for the toxicity impact categories, those recommendations being very much in line with the conclusions of the EC-JRC investigation and the main outcome of the 14th February 2018 stakeholder workshop organised by the Commission.

Using physicochemical and toxicity data available in the REACH-IUCLID database from the European Chemical Agency (ECHA), the OpenFoodTox database from the European Food Safety Authority (EFSA) and from the Pesticide Properties Database (PPDB) from the University of Hertfordshire, new CFs have been calculated for the EF using the USEtox® 2.1 model.

This report describes how freshwater ecotoxicity and human toxicity cancer and non-cancer CFs for application in the EF context have been calculated.

In summary:

— When the required data were available, all substances registered under REACH and/or present in the EFSA database have been included with a CF.

— Substances not in REACH and EFSA database but in the Pesticide Properties database (PPDB) have been added.

— Substances not in REACH / EFSA / PPDB database but in the original USEtox® 2.1 database have been retained.

— Characterisation factors calculated by USEtox® for cationic metals have been retained.

— Human toxicity cancer effect factors are all from the USEtox® 2.1 database have been retained (the cancer required data to run USEtox® 2.1 are not present in REACH/EFSA/PPDB database).

— Human toxicity cancer and non-cancer CFs are based on the USEtox® 2.1 methodology, but using new input data when possible (i.e from REACH, EFSA, PPDB)

— Freshwater ecotoxicity substance hazard values are based on the 20% effect value derived from Species Sensitivity Distribution (SSD) based on chronic ECx equivalent (Pellston workshop agreement - June 2018).

— Since the outcomes of the Pellston workshop require a significant update of the USEtox® 2.1 model for both freshwater ecotoxicity and human toxicity cancer, non-cancer, the EC will decide if and how to take into account the new version of USEtox® including new CFs, once they become available.

— The freshwater ecotoxicity, human cancer and non-cancer impact categories are recommended to be used in EF context, level of recommendation III.

The total number and origin of the new CFs are displayed in the figure 1 below with the following typology:

- **Type 1**: fate, exposure and effect factors are from USEtox® 2.1 original input data.

- **Type 2**: Fate and Exposure have been calculated with REACH physicochemical data, while the effect values are from USEtox® 2.1 original input data.

- **Type 3**: fate, exposure and effect factors have been all calculated with REACH data.

- **Type 4**: fate, exposure and effect factors have been all calculated with EFSA data.

- **Type 5**: fate, exposure and effect factors have been all calculated with PPDB data.

The report is structured as follow:

- **Chapter 2**: Description of the three main databases used to retrieve new physicochemical and toxicity data.

- **Chapter 3**: Selection of the physicochemical data

- **Chapter 4**: Selection and derivation of the substance freshwater ecotoxicity hazard values

- **Chapter 5**: Selection and derivation of the substance human toxicity hazard values

- **Chapter 6**: Calculation of the substance characterisation factors with special consideration regarding organic, inorganic and metal substances as well as for substances reported under a generic names or not yet characterized but being reported in all the EF database. A robustness assessment applied on all CFs is also described.

- **Chapter 7**: A contribution analysis comparing the CFs used by the EF pilots (from 2013-2018) and the new calculated CFs is presented.

- **chapter 8**: Derivation of the normalized factors.

Figure 1: Total number of existing (USEtox® 2.1) and new characterisation factors (EC-JRC-2018) using the REACH, EFSA and PPDB database.
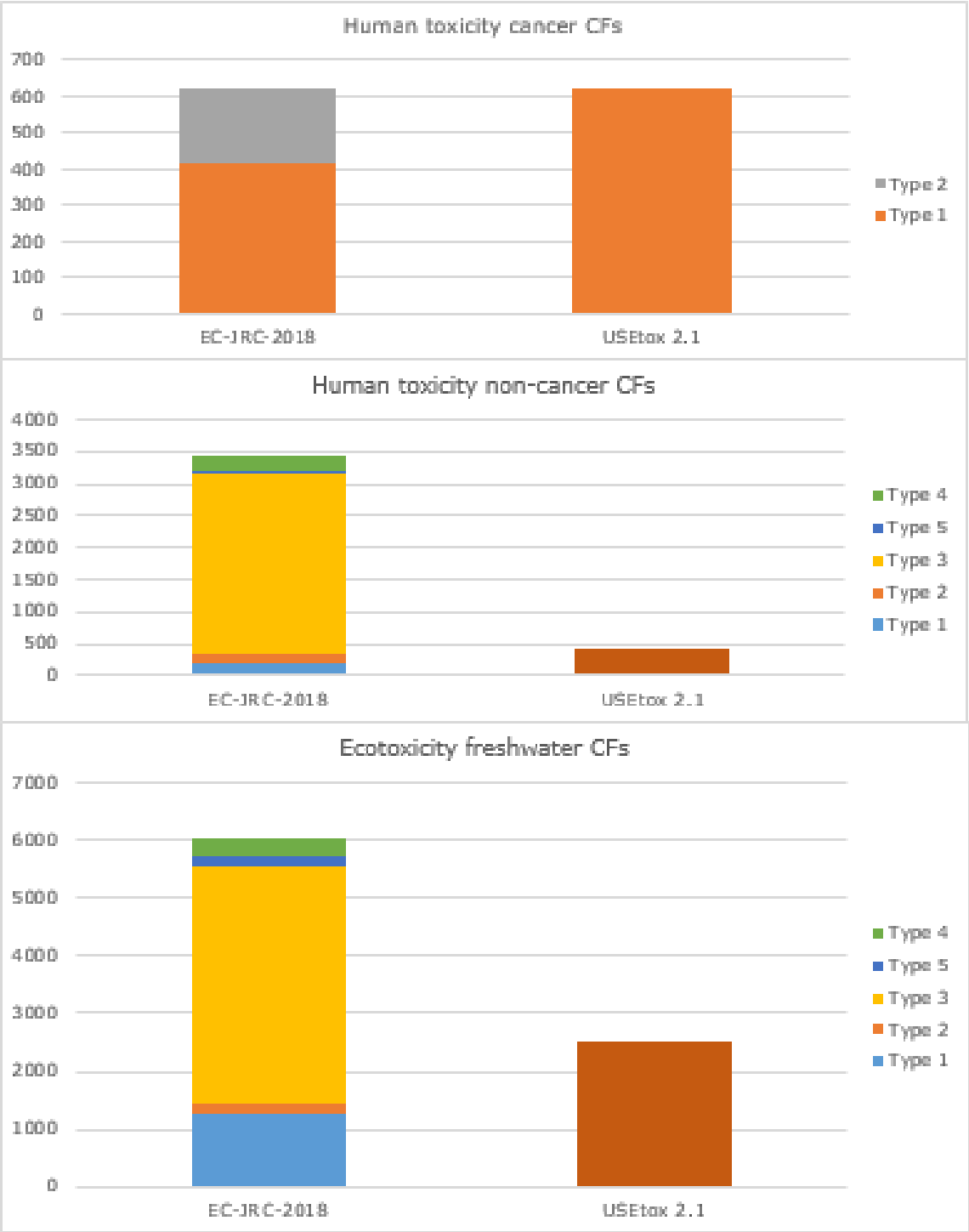
2 Substance properties databases

Three different substances properties databases were used to generate input data and to calculate final substance characterisation factors via the USEtox® 2.1 model.

Two databases come from EU agencies: the REACH-IUCLID database from the European Substance Agency (ECHA) and the OpenFoodTox database from the European Food Safety Agency (EFSA).

The third database is the Pesticide Properties Database (PPDB) developed by the Agriculture & Environment Research Unit (AERU) at the University of Hertfordshire for a variety of end users to support risk assessments and risk management (Lewis et al., 2016).

For substance originally present in the USEtox® database, but not available in one the three databases mentioned above, the USEtox® input data were kept.

2.1 The REACH-IUCLID database

REACH is the European regulation dealing with Registration, Evaluation, Authorisation and restriction of Substances (EC, 2006). Its aims are 1) to guarantee a high level of human health and environment protection from the risks posed by substances, 2) to promote alternative test methods, 3) the free circulation of substances within the European market, 4) to encourage innovation and to enhance competitiveness of the European Union substance industry.

REACH attempts to reach these goals by creating a single system for all substances, replacing all the previous ones; by closing the knowledge gap for more than 30000 existing substances and providing information on both their acute and long-term effects; and by inciting to use and develop safer substances.

As of January 2009, every new substance manufactured or imported above 1 ton in the EU needs to be registered. For the substances already present in the EU market in 2008, a pre-registration step took place between June 2008 and December 2008. For those pre-registered substances, industry must submit a dossier for each substance before:

— December 2010 for substances above 1000 tons, plus the ones above 100 tons if classified N50-53 (very toxic and non-biodegradable), plus the ones > 1 ton if classified CMR (carcinogen, mutagen and repro-toxic).

— June 2013 for substance between 100 and 1000 tons

— June 2018 for substances between 1 and 100 tons

The registration process is a rather complicated task that comprises searching for information, assessing its reliability and relevance, determining the classification & labelling, performing the hazard identification, thinking of additional testing, defining the exposure scenario, calculating the human and environmental risk assessment through the entire life cycle of the substance, completing the substance safety report, communicating through the supply chain.

Thereby, 143000 substances on the European market have been pre-registered. The number and of the type data, short or long-term ecotoxicity, human toxicity is directly linked to the tonnage of the substance marketed or imported to the European Union as it gives an indication of the potential for exposure. To limit the number of experiments and since every substance can only be registered once, REACH legislation strongly encourages the registrants to share all existing data. This is a second strong REACH feature. For this purpose, IUCLID (International Uniform Chemical Information Database) has been developed and allows to collect, store, maintain and exchange relevant data on substance substances (ECHA, 2018a). For the first time, the results of all experiments conducted in industrial laboratories are not kept confidential anymore but become available within the boundaries of the registration process. Furthermore, industries which refuse to share data must justify their choice and can be sanctioned if the justifications are not considered adequate. In addition, if some endpoints or information are missing, the integrated testing

strategy will guide registrants through various alternatives, possibilities, before considering experimental testing. For instance, non-testing data derived from QSAR (Quantitative Structure Activity Relationship) and expert systems can be used to fill data gaps. Read-across and grouping of substances having the same structure and properties are also possible. The data stored in the REACH-IUCLID database are currently used for regulatory purposes for:

— Demonstrating safe use of substances on the EU market for human

— Demonstrating safe use of substances on the EU market for the environment

— For Classification and labelling

— For identification of PBT/vPvB substances (Persistent, Bio-accumulative and Toxic / very Persistent, very Bio-accumulative)

— For identification of SVHC (Substances of Very High Concern)

The REACH regulation requests that all information currently available on a substance must be registered, including the ones of low quality. This was done on purpose to ensure that every single piece of information is taken into account when assessing the safety of substances. However, to avoid that all information is entered into the database without quality discrimination, the regulation has published an extensive guidance documents to help assess the relevance, reliability and adequacy of the information (ECHA, 2017). Detailed guidance is given on information searching strategies and sources of information that may be consulted in the critical first step of assembling all of the available information on a substance, or information that may be useful to inform on the properties of that substance.

All available information that has been gathered on a substance needs to be assessed for its adequacy for classification and labelling, determination of PBT or vPvB status and the derivation of a dose descriptor to be used in the substance safety assessment. The information should be evaluated for its completeness (does the available information meet the information required under REACH) and quality (relevance, reliability and adequacy). Figure 2).

Figure 2: Definition of Relevance, Reliability and Adequacy according to REACH
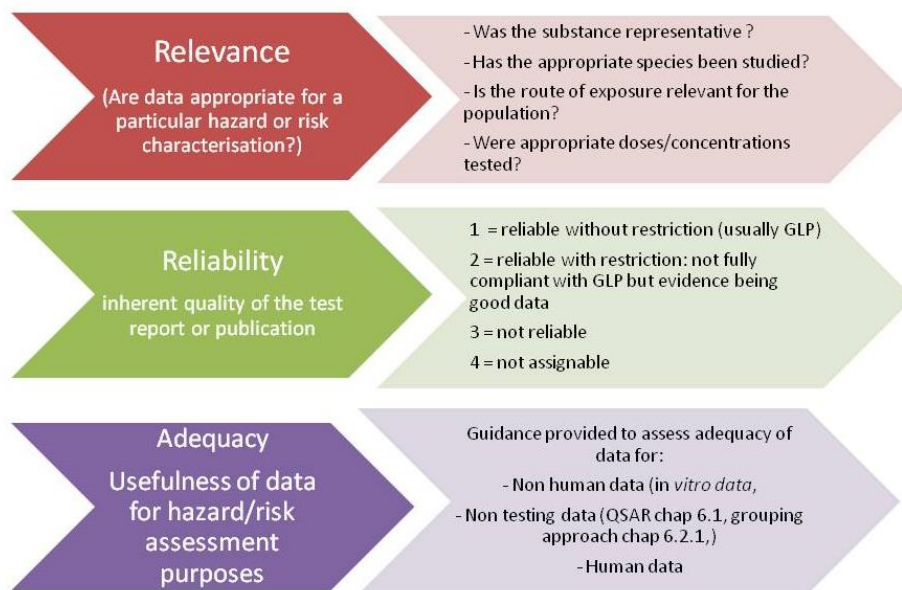
The following definition applies:

— **Relevance**: Relevance is the extent to which data and tests are appropriate for a particular hazard identification or risk characterization.

— **Reliability**: Reliability is the inherent quality of a test report or a publication relating to preferably standardized methodology and the way the experimental procedure and results are described to give evidence of the clarity and plausibility of the findings. It is important to distinguish between reliable methods and reliable information. The Klimisch code is a scoring system for data reliability. The system consists of 4 reliability categories:

- K1: Reliable without restrictions
- K2: Reliable with restrictions
- K3: Not reliable
- K4: Not assignable

— **Adequacy**: Adequacy is the usefulness of the data for hazard and risk assessment purposes.

The REACH guidance document on information requirement also proposes an additional ranking via the use of:

— **key study**: to be used as preference for risk assessment and CLP (Classification, Labelling and Packaging). It represents the most adequate, reliable and relevant for a specific element/endpoint study section. If properly reported, key study may fulfil a REACH information requirement on its own.

— **Supportive study**: is a study that is considered "supportive" of the key study or key studies. A supporting study cannot fulfil a REACH information requirement on its own.

— **Weight of evidence**: The process of considering the strengths and weaknesses of various pieces of information in reaching and supporting a conclusion concerning a property of the substance. This approach always combines a number of individual studies, and none of them can fulfil a REACH information requirement on their own.

Using REACH data for any purpose requires therefore that the users understand the purpose of the regulation and the strategy and rules that registrants had to follow to collect and assess the information.

JRC received from ECHA all the available data from registered dossiers as of May 2015 concerning physicochemical properties, ecotoxicity and human toxicity according to the following IUCLID sections:

— Section 4: Physicochemical properties

- Section 4.6 Vapour pressure
- Section 4.7 Partition coefficient
- Section 4.6 Water solubility
- Section 4.21 Dissociation constant

— Section 5. Fate and pathway

- Section 5.2.1 Biodegradation in water: screening test
- Section 5.2.2 Biodegradation in water and sediment: simulation test
- Section 5.2.3 Biodegradation in soil
- Section 5.4.1 Adsorption / desorption
- Section 5.4.2 Henry's law

— Section 6. Ecotoxicological properties

- Section 6.1 Aquatic toxicity
- 6.1.1 Short term toxicity to fish
- 6.1.2 Long term toxicity to fish

- 6.1.3 Short term toxicity to aquatic invertebrates
- 6.1.4 long term toxicity to aquatic invertebrates
- 6.1.5 Toxicity to algae and cyanobacteria
- 6.1.6 Toxicity to aquatic plants other than algae
- 6.1.7 Toxicity to microorganisms
- 6.1.8 Toxicity to other aquatic vertebrates

— Section 7. Human toxicological properties

- Section 7.5 Repeated dose toxicity
- 7.5.1 Repeated dose oral
- 7.5.2 Repeated dose inhalation

Since May 2015, the number of new registered dossiers for high to medium volume tonnages (annex VII to X) have not increased drastically (see Figure 3). As a matter of fact, 3 registration deadlines have been agreed: 2010 for high volume substances (> 1000 tons) and those classified as toxic for the environment or CMR (Carcinogenic, Mutagenic or toxic to Reproduction); 2013 for substance above 100 tons and June 2018 for low tonnage substances. For those, no significant new toxicity data are expected, as the REACH regulation requires less data than for the higher tonnage bands.

As of 2017, 18835 substances have been registered under REACH with 12494 that have the full set of information required and 6341 that have been registered as intermediate (ECHA, 2018b).

Figure 3: Number of substances registered under the REACH regulation between 2008 and 2017.
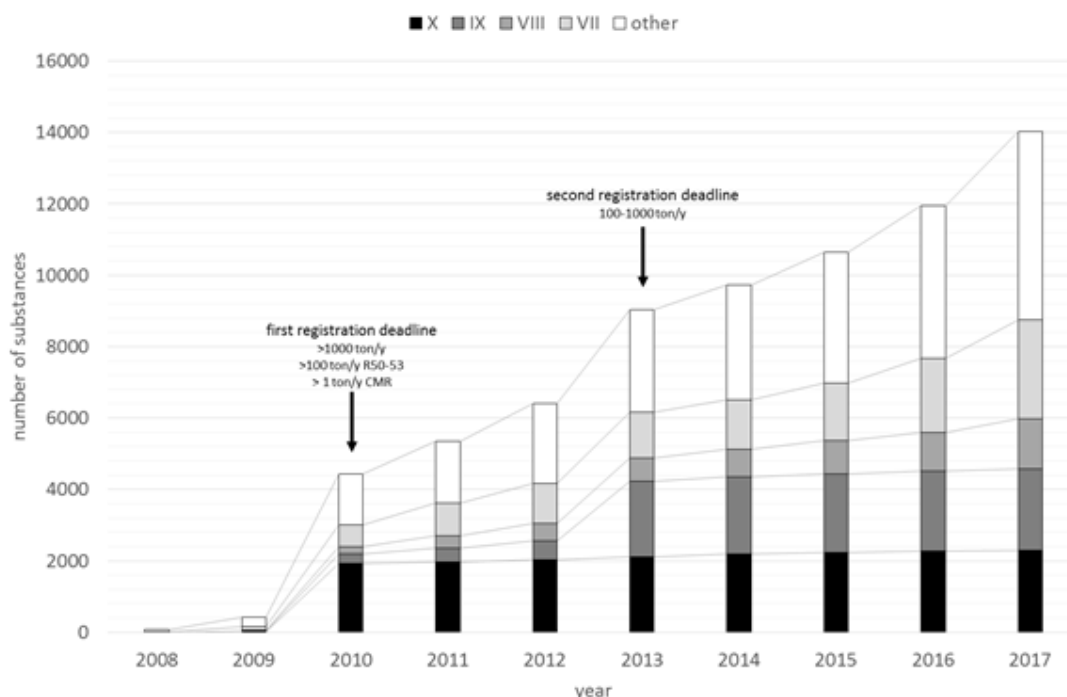
Source: ECHA January 2018

2.2 The OpenFoodTox database

Since its creation in 2002, EFSA (European Food Safety Authority) scientific panels and staff have produced risk assessments for more than 4400 substances in over 1650 scientific opinions, statements and conclusions through the work of its scientists.

EFSA has populated a substance hazards database to hold summary hazard data from EFSA's substance risk assessments in food and feed (Barbaro et al., 2015; Dorne et al., 2017). The database aims at mapping the hazard data as extracted from the EFSA opinions, statements and conclusions, describes the following features. The data repository is updated with all relevant data as collected from EFSA documents (Scientific Opinions, Statements, Conclusions) adopted (and then published) by the Scientific Panels throughout February 2014.

The database aims to hold only summary hazard information from EFSA's previous substance risk assessment on food and feed and <u>not all possible</u> available toxicological data. The database holds information on the substance entity the hazard identification, and the hazard characterisation/risk characterisation. The data repository has been updated with all relevant data as collected from EFSA documents (Scientific Opinions, Statements, Conclusions) that were adopted (and then published) by the Scientific Panels in the past year (up to April 2015).

The data are freely accessible via the EFSA website OpenFoodTox but also accessible via downloadable Excel files (https://www.efsa.europa.eu/en/microstrategy/openfoodtox).

OpenFoodTox is a structured database summarising the outcome of hazard characterisation for human health and – depending on the relevant legislation and intended uses – animal health and the environment

In order to disseminate OpenFoodTox to a wider community, two sets of data can be downloaded:

— Five individual spreadsheets extracted from the EFSA micro strategy tool providing for all compounds: substance characterisation, EFSA outputs, reference points, reference values and genotoxicity

— The full database.

2.3 The PPDB database

The Pesticide Properties Database (PPDB) contains selected quality assessed data on pesticides physicochemical, toxicological, ecotoxicological, human health and other related data (Lewis et al., 2016; PPDB, 2017).

The online version of the PPDB database, launched in 2017, is the result of 20 years effort to collect and format pesticide data to be used freely for conducting substance risk assessment (Lewis et al., 2016). The database contains 2300 actives substances and 700 metabolites. Data have been collected from around 30 different sources, in order to produce a collection as complete as possible.

The use of the database has been acquired by EC-JRC to complement the REACH-IUCLID and OpenFoodTox database in case of missing endpoints (physicochemical properties, human and freshwater toxicity). In order to be consistent, the same procedures followed for the REACH-IUCLID and OpenFoodTox databases were applied to the PPDB dataset.

All the R codes used to deal with PPDB data are reported in the supplementary materials section.

The PPDB brought a significant benefit especially when dealing with pesticide not used or banned in Europe, as those chemicals are out of European legislations and agencies competence. However, those substances can still be used in other parts of the world.

3 Deriving physicochemical properties data from REACH

The R code used to retrieve physicochemical properties from the REACH database and the list of variables available for each parameters are available in the online supplementary material (see annex 1 for list and http://eplca.jrc.ec.europa.eu/LCDN/developerEF.xhtml). The final physicochemical input values used to run the USEtox® 2.1 model are provided with the characterisation factors (see chapter 6) and can be downloaded from the online material.

3.1 Data availability and selection procedure

Fourteen physicochemical and fate properties data are needed to run the USEtox® model, however only seven are mandatory for organic substances and four for inorganic substances (Table 1). Some of the required fields are grouped into the same IUCLID section such as adsorption/desorption coefficients.

Table 1: Physicochemical and fate properties requirements available in the REACH-IUCLID database needed to run the USEtox® model.

USEtox® requirements	Organic substances	Inorganic Substance[1]	In REACH (IUCLID 6 section)
Partition coefficient n-Octanol/Water (Kow)	✓ Mandatory[2]	n/a[3]	✓ (4.7)[4]
Water solubility	✓ Mandatory[5]	✓ Not mandatory	✓ (4.8)
Vapor pressure	✓ Mandatory	✓ Not mandatory	✓ (4.6)
Henry's Law constant	✓ Not mandatory	✓ Not mandatory	✓ (5.4.2)
Biodegradability in water	✓ Mandatory	n/a	✓ (5.2.1)
Biodegradability in air	✓ Mandatory	n/a	No[6]
Biodegradability in soil	✓ Mandatory	n/a	✓ (5.2.3)
Biodegradability in sediment	✓ Mandatory	n/a	✓ (5.2.2)
Partition coefficient organic carbon/water (Koc)	✓ Not mandatory	n/a	✓ (5.4.1)
Partition coefficient dissolved organic and water (KpDOC)	n/a	✓ Mandatory	✓ (5.4.1)
Partition coefficient suspended solid and water (KpSS)	n/a	✓ Mandatory	✓ (5.4.1)

USEtox® requirements	Organic substances	Inorganic Substance[1]	In REACH (IUCLID 6 section)
Partition coefficient suspended sediment and water (Kpsd)	n/a	✓ Mandatory	✓ (5.4.1)
Partition coefficient suspended soil and water (Kpsd)	n/a	✓ Mandatory	✓ (5.4.1)
Acid dissociation constant (pKa)	✓ Not mandatory	n/a	✓ 4.21

[1] In USEtox®, inorganic substances are currently only referring to cationic metals
[2] Mandatory means that this data must be available to run the model.
[3] "n/a" indicates that a parameter is not applicable for this substance group.
[4] IUCLD section where the data are stored.
[5] Not mandatory means that if no specific data point is available, a default value is used by the model.
[6] Instead of biodegradability (does not happen is air), REACH has 'Photo-transformation in air' that could be used (section 5.1.1).

Source: USEtox® 2.1 and ECHA documentation

Table 2 gives an overview to the data received from ECHA in March 2017 (but extracted in May 2015 from the IUCLID database), reporting for each substance properties the number of substances for which a value is available, the number of endpoint study records (ESRs) and number of individual results.

Table 2: Number of substances, endpoint study records (ESRs) and individual results extracted from REACH-IUCLID database for each property.

Physicochemical properties	Number of substances	Number of ESRs	Number of results
Kow	7899	18423	45193
Adsorption/desorption (Koc, Kpss, etc..)	6124	15767	46025
Water solubility	7163	8218	19730
Vapour pressure	7791	16901	24067
Henry's Law constant	1710	2808	3365
Biodegradability (water, sediment, soil)	10809	28359	28546
pKa (only QSAR)	8802		

Each Excel file is organized with each row profiling an endpoint study results and columns reporting variables describing test conditions. Columns containing common information, such as substance identifiers, high and low values, reliability, study type, guidelines, etc, occur in all tables, whilst variables charactering experimental conditions may vary depending on the endpoint. For instance, temperature and pH are reported for many properties, whereas percentage of organic carbon is a highly specific information and thus used only the characterize Koc data. Table 3 illustrates the information reported for each property.

Table 3: Description of information available for each property (see online material for the original list received from ECHA)

Variables:	Kow	Adsorption/desorption	Water solubility	Vapour pressure	Henry's Law constant	Biodegradability – Screening	Biodegradability – Half-life
Dossier UUID	✓	✓	✓	✓	✓	✓	✓
EC number	✓	✓	✓	✓	✓	✓	✓
CAS number	✓	✓	✓	✓	✓	✓	✓
Study Report	✓	✓		✓	✓	✓	✓
Adequacy of study	✓	✓	✓	✓	✓	✓	✓
Study type	✓	✓	✓	✓	✓	✓	✓
Data waiving	✓	✓	✓	✓	✓	✓	✓
Reliability (+ rationale)	✓	✓	✓	✓	✓	✓	✓
Method (+ principles)	✓	✓	✓	✓		✓	✓
GLP compliance	✓	✓	✓	✓	✓	✓	✓
Test material	✓	✓	✓	✓	✓	✓	✓
Guideline (+ qualifier)	✓	✓		✓	✓		✓
Unit of measure	✓	✓	✓	✓	✓		✓
Low value (+ qualifier)	✓	✓	✓	✓	✓		✓
High value (+ qualifier)	✓	✓	✓	✓	✓		✓
Temperature (+ unit)	✓	✓	✓	✓	✓		✓
pH (+ qualifier)	✓		✓				
Remarks	✓	✓	✓	✓			
Type of coefficient	✓	✓					
Matrix/Compartment		✓					✓
% organic carbon		✓					
Oxygen conditions						✓	
Inoculum						✓	✓
Test duration						✓	✓
Qualitative interpretation						✓	
Validity criteria						✓	✓
Test performance						✓	✓
Mineralization rate							✓
Transformation products							✓
Kinetic type							✓
Standard deviation							✓
Sampling time							✓
Degradation parameter							✓

For the majority of fields, a drop down list of predefined options is often paired with a "free text" option, in order to allow the registrants to add details information. This 'free' text option creates some variabilities how the information is recorded according to registrant's skills. Hence, the same information can be retrieved in several columns and, consequently, some pre-processing operations are needed to adjust the database. Those "free text" fields represent one of the principal challenges when dealing with REACH data.

The database presents other important challenges due to data and endpoint variability. In most cases, standard units of measure or endpoints were used for the majority of the data and rules were written on whether and how to use this information. When non-standard unit or endpoint were used, these were often ignored to reduce the complexity of the programming. For instance, more than 70 different unit of measure were found in the database to characterized water solubility results. Moreover, data regarding the biodegradability endpoint are expressed via a qualitative assertion (*i.e.* 'readily biodegradable' or 'inherently biodegradable') and must be converted to a quantitative value to be used in the USEtox® model. The procedure is detailed in the next chapter.

All data treatment and calculation described further in this document have been performed with the RStudio program. Using this software allowed us to build our code in step wise manner until we obtained the desired selection without impacting the structure of the original files (Excel). With the exception of biodegradability, physicochemical properties data are similarly organized and the workflow applied to all of them follows the same structure (Figure 4).

Figure 4: Flow describing the workflow to derive substance physicochemical properties
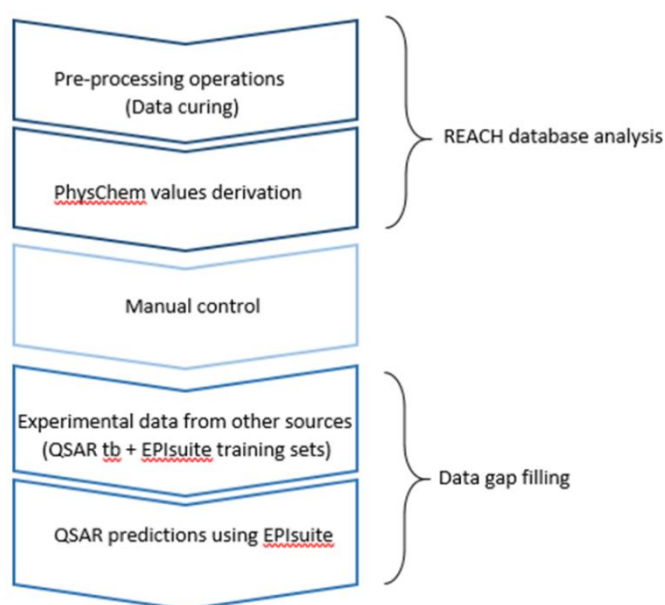
For each parameter, data reported were harmonized using the same unit, the same type of end point and when necessary converted to same temperature (usually 25°C). Duplicate records were also eliminated. Specific treatments applied on each parameter are described in the following sub-chapters.

The following physicochemical and fate properties were retrieved from the REACH data (March 2015) (Table 4).

Table 4: Number of substances per physicochemical and fate test results available in REACH-IUCLID database as May 2015.

Composition and Type of substances		Kow	Pvap25	Sol25	HENRY	Koc	KpSED	KpSUSP	KpSOIL	Biodegr. Screening	Biodegr. Halflife
Mono-constituent	element	27	27	26	5	0	12	10	26	27	26
	inorganic	670	674	621	120	0	191	154	616	668	594
	organic	4108	4090	3929	1134	2951	0	0	0	4131	2057
	organometallic	114	111	84	11	76	0	0	0	108	53
	petroleum	21	20	21	5	14	0	0	0	30	12
Multi-constituent	element	0	0	0	0	0	0	0	0	0	0
	inorganic	47	53	43	4	0	3	2	45	53	40
	organic	564	617	598	93	455	0	0	0	613	309
	organometallic	10	10	8	0	7	0	0	0	10	5
	petroleum	4	3	3	2	3	0	0	0	4	4
	no info/tie	32	27	33	3	23	0	0	0	36	8
UVCB	element	0	0	0	0	0	0	0	0	0	0
	inorganic	166	173	140	18	0	23	20	151	171	150
	organic	1216	1354	1254	244	1185	0	0	0	1350	1067
	organometallic	47	49	45	2	42	0	0	0	48	28
	petroleum	374	375	163	70	368	0	0	0	374	368
	no info/tie	103	109	95	6	89	0	0	0	114	68
no info/tie	element	0	0	0	0	0	0	0	0	0	0
	inorganic	5	4	5	0	0	1	1	4	5	2
	organic	42	40	39	5	35	0	0	0	41	20
	organometallic	1	1	1	0	1	0	0	0	1	0
	petroleum	0	0	0	0	0	0	0	0	0	0
	no info/tie	5	5	5	2	3	0	0	0	5	3
	Total substance	**7556**	**7742**	**7113**	**1724**	**5252**	**230**	**187**	**842**	**7789**	**4814**

## 3.2 Dealing with test results reported as ranges

For some of the physicochemical parameters, test results are expressed with a low and high values. They are associated with descriptors describing the accuracy of the result: '=', 'ca.', '≤', '≥', '<' and '>'.

Strict qualifiers (i.e. '=', 'ca.', '≤', '≥') are considered more reliable, whereas unbounded qualifiers (i.e. '<' and '>') are associated with more uncertainty and generally are linked with limitations of analytical methods.

The following rules were applied to derive one single value per each observation and per substance:

— When no descriptor was provided (or the descriptor 'ca.'), values were considered equivalent to '='.

— When the descriptor was either '>' or '<', the value recorded were used as such, in order not to lose information, but they were assigned to the lowest quality level.

— When value were displayed as ranges (minority of cases), different approaches were used. It was always given priority to 'strict' qualifiers (i.e. '=', 'ca.', '≤', '≥') rather than 'unbounded' qualifiers (i.e. '<' and '>'). On the basis of the qualifiers combination, two possible cases were found:

  ● Combination of two 'strict' or two 'unbounded' qualifiers. The real value is expected to be anywhere between the lower and upper limit of the range, the arithmetic mean is used to derive a unique value (Figure 5 top).

  ● Combination of a 'strict' with an 'unbounded qualifiers'. The real value might be everywhere between the range boundaries, but it is expected to be closer to the limit with the strict qualifier (considered as dominant), therefore only the value associated to it is used, without modifications (Figure 5 bottom).

Figure 5: Rules used to derive one single value per each observation when both a low and high values are available.
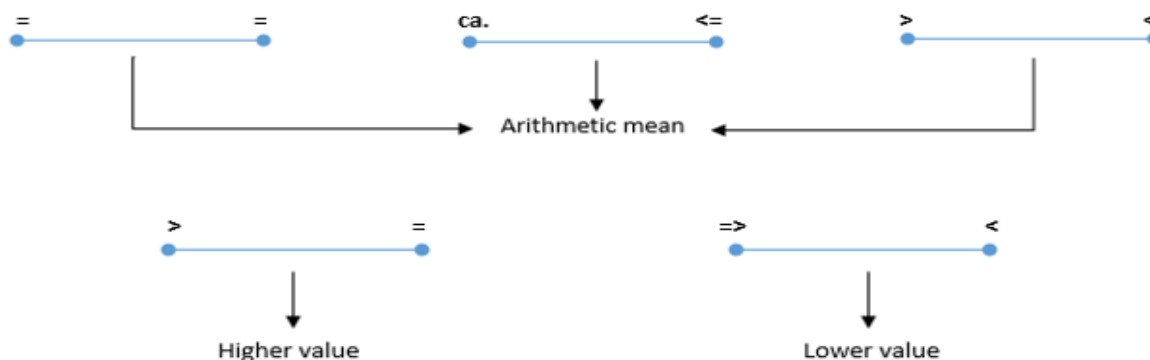
3.3 Attributing a quality score to selected values

Test results were categorized using three quality scores (HIGH, INTERMEDIATE and LOW) using the following criteria (Table 5).

For each quality criteria, the following reasoning was applied:

— **Reliability** is expressed in terms of Klimisch scores, ranging from K1 to K4, where only the first two guarantee proper trustworthiness of the result.

— **Adequacy of study:** No significant differences between "key study", "supporting study" and "weight of evidence" in a data quality perspective were considered.

However, results classified as "disregarded study" or without any description (empty field) were discarded.

— **Study type:** This information was used to prioritize "experimental" results. Other study types include QSAR, calculation, read-across, literature, etc., results with no specified study type were omitted.

— **GLP compliance:** Only results in compliance with Good Laboratory Practices (GLP) were retained in the HIGH quality level; "No" and "Not specified" were not discarded in order not to dramatically reduce the amount of data and to consider also data generated before the introduction of GLP - adopted by OECD in 1981 (OECD, 1981).

— **Qualifier:** A distinction of quality levels was used according to qualifier describing the value: "=" with the HIGH quality score; test values described as "empty", "ca.", ">=" and "<=" with the INTERMEDIATE quality score; whereas unbounded qualifiers are associated with the highest uncertainty and was associated to the LOW quality score.

Table 5: Criteria used to categorize using three quality scores: HIGH, INTERMEDIATE and LOW.

Criteria	HIGH	INTERMEDIATE	LOW
Reliability	K1 + K2	K1 + K2	K1 + K2
Adequacy of study	Key and supporting study + weight of evidence	Key and supporting study + weight of evidence	Key and supporting study + weight of evidence
Type of study	Experimental	All studies	All studies
Qualifier	=	">=" and "<="	">" and "<"
GLP compliance	Yes	Yes, No, Not specified	Yes, No, Not specified

For each substance and each physicochemical parameter, if several test results were available, only the highest quality was retained to be used in the USEtox® model. For instance, assuming that several results meeting HIGH and INTERMEDIATE quality levels criteria were reported for one substance and one parameter, only the outcomes labelled with the HIGH level score were retained.

When for a substance, more results associated to the same quality score were available, the geometric mean (GM) was calculated to generate a unique value. In addition, geometric coefficient of variation (GCV%) was calculated. GM and GCV% were adopted in order to reflect the log distribution of the physicochemical properties data.

Finally, when a 'key value' was reported in the endpoint study summary section of the REACH Chemical Safety Assessment (CSA), this value was always selected as the final value to be used with the USEtox® model, as they represent the best value according to registrants expertise.

3.4 Selection procedure

3.4.1 Partitioning coefficient between n-octanol and water

The partitioning coefficient between n-octanol and water (Kow) is one of the key parameter used in fate modeling to estimate the distribution of substance between the water phase and organic carbon present in the environment (particles, sediment, biota, etc.). It expresses the substance affinity with hydrophilic or hydrophobic phases. In toxicology, it represents the tendency of a compound to pass through biological membranes as well as to accumulate in tissue (fat). Mathematically, it is calculated as the ratio of concentration of a substance mixed into two solvents: n-octanol and water.

For 7899 substances, having a Kow value reported in REACH-IUCLID, 21515 individual test observations were available.

For 792 substances for which a Chemical Safety Assessment was available, the proposed "key value for safety assessment" was selected as the final value to be used with USEtox®.

Table 6 summarizes the characteristics of test records registered in REACH-IUCLID database. In particular, the different endpoint, analytical methods, adequacy of the study, reliability, low and high values descriptors are shown.

Table 6: Number of n-octanol water partition coefficient recorded in the REACH-IUCLID database (total = 21515) according to test methods, study adequacy, reliability, and if a low and high values were reported.

Methods		Endpoint	
Shake flask method	3728	Pow	1461
HPLC method	6160	LogPow	19919
Generator column method	20	empty	135
Slow stirring method	150		
Not specified	11457		

Adequacy		Reliability		Low value		High value	
Key study	12164	K1	7759	>	1886	<	1173
Supporting study	5211	K2	11408	>=	472	<=	733
Weight of evidence	2109	K3	312	ca.	1429	ca.	75
Disregarded study	112	K4	1043	empty	17728	empty	19534
empty	1919	empty	993				

In the REACH-IUCLID database, the n-octanol/water partition coefficient has been reported either as Pow or logPow (Pow equivalent to Kow; as P stands for partition and K being the equilibrium constant). Values registered as logPow or empty were converted in their non-logarithmic form. Data with no endpoint information (blanks) were discarded. Empty were considered to be equivalent to logPow. Any eventual odd-looking values generated during conversion (too high or too low) were detected and discarded when evaluating method sensitivity. Several analytical methods were developed for Kow determination, in order to be applied on a wide range of substances with different properties (e.g. gas, liquid, solid). The domain of application of each analytical method was considered to define the validity of the results, meaning that values below or above certain limits are deemed unreliable.

Overall, three parameters verified:

— Temperature: accepted values within the range $10°C \leq T \leq 30°C$;

— Acidity: accepted values within the range $5 \leq pH \leq 9$;

— Kow value, according to the analytical method:

  ● Shake flask method: $0.01 \leq Kow \leq 10000$;
  ● HPLC method: $1 \leq Kow \leq 1000000$;
  ● Generator column method: $10 \leq Kow \leq 1000000$;
  ● Slow stirring method: $0.01 \leq Kow \leq 100000000$;
  ● Not specified method: $01 \leq Kow \leq 10000$.

Only results satisfying these requirements were accepted, the rest of the database was omitted. This step can help in avoiding outliers: very large or negative Kow values (not realistic) will be judged as outside of method sensitivity and discarded.

3.4.2 Water solubility

Water solubility (Sol25) refers to the ability of a substance to dissolve in water. This parameter is strongly temperature-dependent. In USEtox®, it is expressed as concentration in mg/L at the temperature of 25°C.

Table 7 summarizes the characteristics of test records registered in REACH-IUCLID database. In particular, the different endpoint, analytical methods, adequacy of the study, reliability, low and high values descriptors are shown.

Table 7: Number of water solubility values in the REACH-IUCLID database (total 18664 records) according to test methods, study adequacy, reliability, and if a low and high values were reported.

Methods		Low value		High value	
Column elution method	1494	>	841	<	2476
Flask method	5733	>=	428	<=	677
Not specified	11437	ca.	1352	ca.	93
		empty	16043	empty	15418

Adequacy		Reliability	
Key study	10568	K1	6560
Supporting study	4564	K2	9543
Weight of evidence	1706	K3	298
Disregarded study	205	K4	1321
empty	1621	empty	942

For 7163 substances, having a water solubility value reported in REACH-IUCLID, 18664 individual test observations were available.

For 756 substances for which a Chemical Safety Assessment was available, the proposed "key value for safety assessment" was selected as the final value to be used with USEtox®.

In REACH-IUCLID database, several unit of measure were used (e.g. mg/L, ppb, kg/m3, vol%, etc.), therefore a conversion was necessary. Ambiguous unit of measure, such as "parts per water" and "%", were not converted. Two main analytical methods were developed for water solubility determination: the "column elution method" and the "Flask method" for highly and scarcely soluble compounds, respectively. The domain of application of both methods was considered to define the validity of the results, meaning that values below or above certain limits are deemed unreliable.

Overall, three parameters verified:

— Temperature: accepted values within the range 10°C≤T≤30°C;

— Acidity: accepted values within the range 5≤pH≤9;

— Sol25 value, according to the analytical method:

  • Column elution method: Sol25 ≥ 10 mg/L;
  • Flask method: Sol25 ≤ 10 mg/L.

Only results satisfying these requirements were accepted. For few cases, the reported value was the limit of analytical determination, being the real solubility value outside of the instrument sensitivity, in those situations the values were assigned to the LOW quality level.

USEtox® model requires water solubility precisely at 25°C, because this value is highly temperature-dependent and its internal algorithm convert this parameter to the equivalent at the temperature of the different media compartments. The conversion is based on a function of state, the following equation is used:

Equation 1:

$$Sol_{Tf} = Sol_{Ti} \cdot exp\left(\frac{H_{diss}}{8.314} \cdot \left(\frac{1}{Ti} - \frac{1}{Tf}\right)\right)$$

Where Ti is the initial temperature in K (before conversion), Tf is the final temperature in K (after conversion), Hdiss is the enthalpy of dissolution which is equal to 10000 J mol-1, SolTf and SolTi are the solubility values at the temperature of Ti and Tf, respectively.

This equation is applied to convert all solubility values to the temperature of 25°C (T_f=298K), however values with a temperature out of the range of 10-30°C (T_i≤283K or T_i≥303K) were excluded in order not to consider values obtained in experimental conditions exceeding method reliability.

The same equation is run in the USEtox® model to generate solubility values typical for each media, in this case T_i=298 K and T_f is the compartment temperature (e.g. 285K for continental landscape).

3.4.3 Vapor pressure

Vapor pressure (Pvap25) represents the tendency of substances to escape from condensed phases (solid or liquid) to vapor form. This parameter is strongly temperature-dependent. In USEtox®, it is expressed in Pa (Pascal, 1Pa = 1kg m-1 s-2) at the temperature of 25°C.

Table 8 summarizes the characteristics of test records registered in REACH-IUCLID database. In particular, the different endpoint, analytical methods, adequacy of the study, reliability, low and high values descriptors are shown.

For 7791 substances, having a vapour pressure value reported in REACH-IUCLID, 21813 individual test observations were available.

For 575 substances for which a Chemical Safety Assessment was available, the proposed "key value for safety assessment" was selected as the final value to be used with USEtox®.

Table 8: Number of vapour pressure values in the REACH-IUCLID database (total 21813 records) according to test methods, study adequacy, reliability, and if a low and high values were reported

Methods		Low value		High value	
Dynamic method	1959	>	137	<	2447
Static method	2845	>=	155	<=	504
Isoteniscope method	385	ca.	1332	ca.	5
Gas saturation method	789	empty	20189	empty	18857

Spinning rotor method	55				
Effusion method	3105				
Not specified	12675				

Adequacy		Reliability	
Key study	11887	K1	6063
Supporting study	5189	K2	12380
Weight of evidence	2190	K3	232
Disregarded study	147	K4	1691
empty	2400	empty	1447

In REACH-IUCLID database, several unit of measure were used (e.g. Pa, bar, atm, mm Hg, psi, etc.), therefore a conversion was necessary. Several analytical methods were developed for vapour pressure determination, based on different physical laws. The domain of application of each analytical method was considered to define the validity of the results, meaning that values below or above certain limits are deemed unreliable.

— Overall, two parameters verified:

— Temperature: accepted values within the range 10°C≤T≤30°C;

— Vap25 value, according to the analytical method:

- Dynamic method: 1000≤Pa≤100000;
- Static method: 10≤Pa≤100000;
- Isoteniscope method: 100≤Pa≤100000;
- Gas saturation method: 0.00001≤Pa≤1000:
- Spinning rotor method: 0.0001≤Pa≤0.5;
- Effusion method: 0.001≤Pa≤1:
- Not specified method: 0.00001≤Pa≤100000.

Only results satisfying these requirements were accepted.

Analogously to water solubility, vapour pressure is strongly temperature-dependent and it must be converted to 25°C. The following equation was used:

Equation 2:

$$Vap_{Tf} = Vap_{Ti} \cdot exp\left(\frac{H_{vap}}{8.314} \cdot \left(\frac{1}{Ti} - \frac{1}{Tf}\right)\right)$$

Where T_i is the initial temperature in K (before conversion), T_f is the final temperature in K (after conversion), H_{vap} is the enthalpy of vaporisation which is equal to 50000 J mol^{-1}, Vap_{Tf} and Vap_{Ti} are the pressure values at the temperature of T_i and T_f, respectively. This equation is applied to convert all pressure values to the temperature of 25°C (T_f =298K), however values with a temperature out of the range of 10-30°C (T_i ≤283K or T_i ≥303K) were excluded in order not to consider values obtained in experimental conditions exceeding method reliability.

3.4.4 Henry's law constant

Henry's law constant (KH25C) is the measure of the concentration of a substance in air over its concentration in water. It reflects the relative volatility of a particular substance and it is a pivotal property when modelling fate and transport of chemicals. In USEtox®, it is expressed in Pa m^3 mol^{-1}.

Table 9 summarises the characteristics of test records registered in REACH-IUCLID database. In particular, the different endpoint, analytical methods, adequacy of the study, reliability, low and high values descriptors are shown.

For 1710 substances, having a Henry's law constant value reported in REACH-IUCLID, 2886 individual test observations were available.

For 170 substances for which a Chemical Safety Assessment was available, the proposed "key value for safety assessment" was selected as the final value to be used with USEtox®.

After conversion, only temperature outside the range 10°C≤T≤30°C were discarded.

Table 9: Number of Henry's law constant records in the REACH-IUCLID database according to test methods, study adequacy, reliability, and if a low and high values were reported

Adequacy		Reliability		Low value		High value	
Key study	1867	K1	53	>	11	<	78
Supporting study	523	K2	2627	>=	11	<=	13
Weight of evidence	257	K3	18	ca.	94	ca.	0
Disregarded study	23	K4	149	empty	2770	empty	2795
empty	216	empty	39				

3.4.5 Sorption coefficients

Sorption coefficients reflects the tendency of substances to be absorbed/desorbed on different environmental matrices, they are expressed in terms of partitioning coefficients. All partitioning coefficients necessary to run USEtox® are recorded in section 4.5.1 of IUCLID. Proper curing operations are necessary to extract coefficient for suspended soil/water partition (KpSS), sediment/water partition (KpSED), soil/water partition (KpSOIL) and organic carbon/water partition (Koc).

In order to divide the sorption data in the correct coefficients, an initial distinction between organic and inorganic substance was performed. Organics were always associated to Koc. Then, for a further refinement of the selection, inorganic substances were retrieved based on specific key words describing the matrix type: suspended soil, sediment or soil. However, it occurred that conflicting information were reported in different column for the same result. Such disagreements were always solved in favor of suspended solids partition coefficient, being the most common and simple experiment.

For 6124 substances, having a sorption partition coefficient value reported in REACH-IUCLID, 13971, 2383, 5021 and 4300 individual test observations were available for Koc, KpSED, KpSUSP and KpSOIL, respectively (Table 10).

For 478, 16, 13 and 17 substances, for which a Chemical Safety Assessment was available for Koc, KpSED, KpSUSP and KpSOIL, respectively, the proposed "key value for safety assessment" were selected as the final value to be used with USEtox®.

Temperature is the only parameter considered to define valid experimental conditions: results within the temperature interval of 10°C≤T≤30°C were accepted.

Several endpoints are reported to describe the tendency of substance toward the sorption process, mainly Koc, Kp, Kd and their logarithmic forms; these endpoint were converted using the following equation, describing the relation between Koc and Kp:

Equation 3:

$$Koc = \frac{Kp}{weight\ fraction\ of\ organic\ carbon}$$

The weight fraction of organic carbon was retrieved from the appropriate column of the database. However, this information was scarcely available and default values from the "Guidance on information requirements and substance safety assessment. Chapter R.16: Environmental Exposure Estimation" were used (ECHA, 2017). These values are collected in Table 11. When no indication about matrix or compartment it was assumed to be soil.

Table 10: Number of adsorption / desorption values in the REACH-IUCLID database according to test methods, study adequacy, reliability, and if a low and high values were reported

	Adequacy		**Reliability**		**Low value**		**High value**	
KOC	Key study	7798	K1	2990	>	1001	<	969
	Supporting study	2185	K2	9949	>=	347	<=	341
	Weight of evidence	3070	K3	108	ca.	296	ca.	44
	Disregarded study	60	K4	358	empty	12227	empty	12517
	empty	758	empty	466				
	Total	13871		13871		13871		13871
	Adequacy		**Reliability**		**Low value**		**High value**	
KpSED	Key study	287	K1	1	>	129	<	123
	Supporting study	458	K2	2166	>=	119	<=	119
	Weight of evidence	1594	K3	103	ca.	52	ca.	8
	Disregarded study	24	K4	113	empty	2083	empty	2133
	empty	20	empty	0				
	Total	2383		2383		2383		2383
	Adequacy		**Reliability**		**Low value**		**High value**	
KpSUSP	Key study	321	K1	1	>	56	<	95
	Supporting study	967	K2	4947	>=	1	<=	1
	Weight of evidence	3729	K3	45	ca.	16	ca.	0
	Disregarded study	4	K4	28	empty	4948	empty	4925
	empty	0	empty	0				
	Total	5021		5021		5021		5021
	Adequacy		**Reliability**		**Low value**		**High value**	
KpSOIL	Key study	1676	K1	261	>	226	<	118
	Supporting study	1279	K2	3659	>=	34	<=	58
	Weight of evidence	1030	K3	320	ca.	77	ca.	0
	Disregarded study	210	K4	58	empty	3963	empty	4124
	empty	105	empty	2				
	Total	4300		4300		4300		4300

Table 11: Weight fraction of organic carbon

Matrix/compartment	Weight fraction of organic carbon
Suspended soil	0.01
Sediment	0.05
Soil	0.02

Source: EU Technical Guidance Document (ECB, 2003)

3.4.6 Degradation rate in water, sediment and soil

USEtox® requires degradation in water, sediment and soil and air. Degradation in the first 3 compartments are usually related to bio-degradation (via microorganisms), while degradation in air is usually due to photo-transformation.

In USEtox®, the Biowin3 model was used to convert the ultimate biodegradation probability in half-lives for all substances in the database (Boethling et al., 1994). Division factors of 1:2:9 were used to extrapolate biodegradation rates for water, soil and sediment compartments respectively, as suggested in EPISuite™ (US-EPA, 2012).

REACH has biodegradation data in wastewater treatment, surface water, sediment and soil. Hydrolyse data are also available. For the air compartment, the photo-transformation data could be used, but this remains to be explored.

The biodegradation data in REACH are recorded in 3 different sections:

— Section 5.2.1 Biodegradation is water: screening tests

— Section 5.2.2 Biodegradation in water and sediment

— Section 5.2.3 Biodegradation is soil

Two approaches were followed to estimate substance biodegradability in surface water, sediment and soil.

3.4.6.1 Using biodegradation half-life from screening tests

The readily biodegradability test results are reported in the 5.2.1 section and are used to discriminate likelihood for substances to be biodegradable in the environment. Seven predefined options are available in IUCLID (Table 12).

Table 12: Total number of results per pre-defined options in section 5.2.1 of IUCLID

IUCLID predefined results interpretation	Number of ESRs	Harmonized biodegradation category
Readily biodegradable	5917	Readily biodegradable
Readily biodegradable, but failing 10-days window	596	Biodegradable, failing 10-days
Inherently biodegradable	1904	Inherently biodegradable Inherently biodegradable Inherently biodegradable
Inherently biodegradable, fulfilling specific criteria	44	
Inherently biodegradable, not fulfilling specific criteria	165	
Not inherently biodegradable	751	Not readily biodegradable Not readily biodegradable
Under test conditions, no biodegradation observed	2852	
Others	8613	Information provided in another column

Screening tests are more widely performed and available from the REACH-IUCLD database. Reliable qualitative interpretations were retrieved for 5276 substances.

Due to strict test conditions (high substance concentration – 10 or 20 ppm – and very little amount of bacteria), it is usually considered that if a substance biodegrade under these conditions, it is more than likely that it will also degrade fast in the real environment (and even faster in wastewater treatment plants (WWTP)). The results of this test (the most frequently available test in REACH dossiers) is used in the risk assessment model EUSES (European Union System for the Evaluation of Substances) to define degradation constant in WWTP, surface water, sediment and soil using default rate constant (Table 13).

However, the Technical Guidance Document (TGD) proposes a rate constant for four categories (Table 13) while seven are available in the IUCLID section. We therefore grouped some categories from the IUCLID section to fit the categories proposed by the TGD (see last column of Table 12)

When several screening tests were available for the same substance reporting different outcome, if at least one test report "readily biodegradable" the substance was classified as "readily biodegradable"; using the same approach, results indicating faster degradation were always considered more relevant when paired with assessments expressing slower or zero degradation. As consequence, "Not readily biodegradable" was assigned only when all interpretations available agreed on its persistence.

3.4.6.2 Using biodegradation half-life simulation tests

These tests attempt to simulate degradation in a specific environment by use of indigenous biomass, media, relevant solids (i.e. soil, sediment, activated sludge or other surfaces) to allow sorption of the substance, and a typical temperature that represents the particular environment. These are available in the IUCLID section 5.2.2 and 5.2.3.

A representative and low concentration of test substance is used in tests designed to determine the biodegradation rate constant whereas higher concentrations for analytical reasons are normally used for identification and quantification of major transformation products (ECHA, 2015).

Experimental half-life values in water and sediments were extracted for 180 and 164 substances, respectively, from section 5.2.2. Half-life values were then converted to degradation rates using the following equation:

Equation 4:

$$k_{deg} = \frac{\ln 2}{HL \cdot 86400}$$

Where *kdeg* is the degradation rate expressed in s-1, *HL* the half-life in day, and 86400 s/d is the days to second conversion factor.

3.4.6.3 Assigning half-life

A default degradation rate was assigned to each harmonized biodegradation category from then screening tests. Default values reflect those reported in the TGD Part II 2.3.6.5 (ECB, 2003), except the one for "not readily biodegradable" category, where the rate of 0 s-1 of the TGD (no degradation, typical of metals) was replace by a degradation rate of 4.45E-8 s-1 from the Biowin3 'recalcitrant' category (US-EPA, 2012).

Table 13 reports the suggested default values for each category, as suggested by TGD and USEtox® interpretations of the Biowin3 results. In addition, the geometric mean of the experimental half-life value was compared to verify the quality of the TGD information. "Readily biodegradable" category experimental mean value appears to be coherent with the one from the TGD; unfortunately, no conclusions can be drawn for the remaining categories because of the very little of substances representing them.

As suggested in EPISuite™ (US-EPA, 2018), division factors of 1:2:9 were applied to extrapolate biodegradation rates for water, sediment and soil, respectively; the quality level associated to KdegW was assigned also to degradation rates in sediment and soil.

Table 13: Equivalence between biodegradation categories and degradation rate constant from the EU Technical Guidance Document.

Biodegradation category	Degradation rate constant (s-1)	Half-life (in days)
Readily biodegradable	5.38E-07	15
Biodegradable, failing 10-days	1.6E-07	50
Inherently biodegradable	5.35E-08	150
Not readily biodegradable	0	∞

3.4.7 Data gap filling procedure

After exploring the REACH-IUCLID database, many substances were lacking a value for all the physicochemical properties. In order to 'guarantee' a value for each parameter, a data gap filling procedure was performed using the OECD QSAR toolbox (Wegmann et al., 2009) and the EPIsuite™ estimation software (US-EPA, 2018). The Pesticide Property Database (Lewis et al., 2016) was also consulted in case the other approaches failed.

The first includes useful tools for profiling substances, searching properties in existing databases and predicting values using read-across. However, QSAR toolbox cannot run

batch read-across and it cannot provide uniformity and reproducibility. In fact, user decisions and expertise on read-across are dominant. For these reasons, QSAR toolbox was used only to collect experimental data from existing databases.

On the contrary, EPISuite™ can provide predictions for several properties. Moreover, experimental data used to train the models can be extracted. Unfortunately, EPISuite™ does not provide a strict definition of its AD (Applicability Domain) however, it suggests it is reasonable to consider substances whose molecular weight (MW) falls within the MW range of the training set to be more reliable.

Following this consideration, the following order of preference was adopted:

1. Experimental data from training sets

2. EPIsuite™ prediction inside AD

3. EPIsuite™ prediction outside AD.

Kp values were excluded from the data gap filling procedure as neither QSAR toolbox nor EPIsuite™ contain this type of data.

Brief description of EPIsuite™ QSAR models used for data gap filling purposes:

— **Kow**. KowWIN model (Meylan & Howard, 1995) uses a "fragment constant" methodology to predict LogKow. Quality is expressed through the correlation coefficient: $R2=0.943$ on the validation set for substances within the AD and $R2=0.879$ for substances outside AD.

— **Water solubility**. WATERNET (Meylan et al., 1996) estimates water solubility at 25°C ($R2=0.815$); it uses a "fragment constant" methodology, based on large training and validation sets (1128 and 4636 respectively).

— **Vapour pressure**. Vapour pressure at 25°C (mm Hg) was estimated by the Modified Grain method (for solids) (Neely, 1985) and by the average between Antoine (Lyman et al., 1990) and Modified Grain methods (for gases and liquids), using the MpBpVpWIN model in EPIsuite™. The training set was built using experimental values at different temperatures. Therefore, few data curing operations were performed, by discarding results outside the validity range of 10°C≤T≤30°C and converting values to 25°C.

— **Henry's Law**. Experimental data was searched in the HenryWIN training set (Meylan & Howard, 1991). When experimental information was lacking, Henry's law constant was calculating using equation 5 (see below).

— **Koc**. The partitioning was estimated using the Molecular Connectivity Index (MCI) approach included in the KocWIN model in EPIsuite™ (Meylan et al., 1992). It predicts Koc on the basis of the molecular structure ($R2=0.778$).

— **Degradation rate in water**. The ultimate degradation probability estimated by Biowin3 (Boethling et al., 1994) was converted in half lives and rate constants as suggested in USEtox® documentation (Table 14). Division factors of 1:2:9 were then applied to extrapolate rates for water, sediment and soil. Last, being Biowin3 trained only on organic substances, all predictions for inorganic compound were discarded and replaced with an arbitrary extremely low value of 1E-20 s-1.

Equation 5:

$$KH25C = \frac{Pvap25 \cdot MW}{Sol25}$$

Where Pvap25, MW and Sol25 are vapour pressure at 25°C (Pa), molecular weight (g/mol) and water solubility at 25°C (mg/L). EPIsuite™ models generates estimations from the molecular structure. It can be entered using either a SMILES (Simplified Molecular Input Line Entry System) or a CAS number, provided that it is included in the internal SMILESCAS

database (known CAS/SMILES associations) (US-EPA, 2018). SMILESs were retrieved using the OECD QSAR toolbox. However, for many substances, the SMILES was reported with an invalid notation that cannot be recognized by EPIsuite™. The data gap filling process failed for substances with an invalid SMILES notation, a not recognized CAS and no information in the OECD QSAR toolbox.

Table 14: USEtox® half life and degradation rates for Biowin3 estimations.

Biowin3 output	Assigned half-life (d)	Degradation rate (s-1)
Hours	0.17	4.7E-05
Hours to days	1.25	6.4E-06
Days	2.33	3.4E-06
Days to weeks	8.67	9.3E-07
Weeks	15	5.3E-07
Weeks to months	37.5	2.1E-07
Months	60	1.3E-07
Recalcitrant	180	4.5E-08

3.5 Other physicochemical properties not available in REACH

In addition to the parameters retrieved, USEtox® model requires data for bioaccumulation factor for fish (BAFfish), degradation rate in air (kdegA) and acid dissociation constant (pKa).

BAFfish and kdegA were retrieved from EPIsuite™, following the procedure described in USEtox® documentation (Fantke et al., 2017).

Dissociation constant is reported in USEtox® with three parameters: pKaChemClass, pKa.gain and pKa.loss. In details, pKaChemClass indicates the nature of the organic substance ("acid", "base", "amphoter" or "neutral"), pKa.loss and pKa.gain represent the equilibrium constant of the dissociation reactions of the acid and the base's conjugated acid, respectively. These were retrieved using ADMET® predictor (Simulations-Plus, 2016) (SPARC® 6.0 was used in USEtox®). From ADMET® output, the value of the first acid dissociation was used as pKa.loss and the value of the first basic dissociation as pKa.gain. Finally, pKaChemClass was assigned based on the presence/absence of pKa.loss and pKa.gain: "acid" if only pKa.loss was predicted, "base" is only pKa.gain was available, "amphoter" if both are reported and "neutral" if none are recorded

3.6 Combining data from different sources

Considering the characteristics of the quality levels from REACH-IUCLID database and the nature of the data gap filling tools, the following order assessing the quality and reliability of data, from best to worst, was used to combine outcomes from different sources. Final values were prioritized according to the following order:

— Key value for safety assessment (REACH)

— High reliability (REACH)

— Experimental data from OECD QSAR toolbox or EPIsuite™ training sets.

— Intermediate reliability (REACH)

— EPIsuite™ prediction inside AD

— Low reliability (REACH)

— EPIsuite™ prediction outside AD

— PPDB

— USEtox® 2.1 input datasheet

For Henry's law constant no EPIsuite™ estimations were done, these are substituted by the ratio between vapour pressure and water solubility.

Last, for a certain number of substances listed in USEtox® input data sheet it was not possible to retrieve new parameters from REACH-IUCLID. Either because the substance was not registered in REACH or data didn't meet our quality requirements. In this case, the original USEtox® physicochemical properties were kept.

3.7 Final physicochemical properties to be used with the USEtox® 2.1 model

Physicochemical properties data were retrieved for 10270 substances; for which it will be possible to generate fate and exposure factors with the USEtox® model. Nine properties (Kow, Pvap25, Sol25, KH25C, KdegW, Koc, KpSS, KpSED and KpSOIL) were retrieved from REACH database, when available. A selection was performed in order to guarantee the extraction only of the most reliable information; a quality level was assigned to each result in order to assess its quality (table 15).

Table 15: Total number of values per physicochemical parameter extracted from the REACH-IUCLID database (May 2015) per quality scores.

Parameter	Total	Key values from CSA*	High quality	Intermediate quality	Low quality
Octanol-water partition (Kow)	2011	772	574	659	6
Vapor pressure (Pvap25)	1648	563	412	666	7
Water solubility	2175	735	572	829	39
Henry's law constant	828	168	7	653	-
Degradation in water	3247	337	1756	1154	-
Organic Carbon adsorption coefficient (Koc)	1972	468	314	1178	12
Suspended Solid – water partition coefficient	115	13	-	101	1
Sediment – water partition coefficient	139	16	3	120	1
Soil – water partition coefficient	185	17	24	142	2

*Reach Chemical Safety Assessment (CSA) dossier

Being the best data available, all KEY VALUES FOR SAFETY ASESSMENT were retained; the little number of substance with LOW quality, for all properties, suggests that quality score system guarantee a proper skimming of data saving only the most reliable results.

In addition, OECD QSAR toolbox, EPIsuite™ and PPDB were adopted for data gap filling purposes. Histogram in figure 6 represents the amount of results collected from the different sources and their quality level for the main six physicochemical parameters. KpSS, KdSED and KpSOIL were not considered because they are excluded from the data gap filling procedure; pKa, KdegA and BAFfish were not considered since they were retrieved only from one tool.

Figure 6: Origin and quality assessment of the 10270 physicochemical data used to calculate fate factor with the USEtox® model.
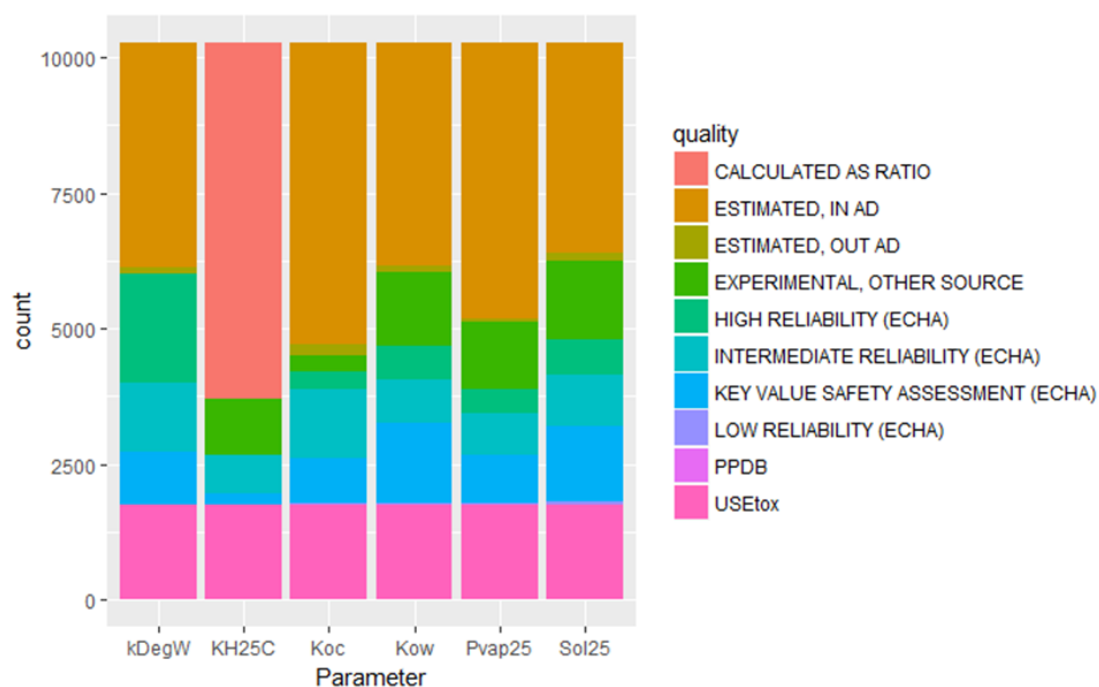
REACH-IUCLID provides a large load of data, in different amount for each property according to the availability of their results, supported by EPIsuite™ estimation models which contributes with the majority of information.

For a significant number of substances it was not possible to provide new experimental data and USEtox® input values were maintained. Many reasons identified:

— around 200 substances in USEtox® are registered as intermediate in REACH and for those substances, no registration is needed.

— some substances are registered in Annex III (those are substances between 1 to 10 tons that, if meeting Annex III criteria, need to provide full Annex VII information). If no registration a dossier is available, it is likely that the registrant may have decided not to use this substance anymore. It is therefore unlikely that new data will become available.

— Some substances are not used in Europe and therefore not required to be registered in REACH.

Scatter plots from figure 7 to figure 12 show a comparison between the new values proposed in this report and those listed in USEtox® 2.1 (which are all from the EPIsuite™) for six physicochemical properties. In each figure, the top graph shows the source of the data and the bottom graph show the ratio (expressed in log) between USEtox® 2.1 and new EC-JRC-2018 values (with in parenthesis the number of observation falling in each bin).

Figure 7: Comparison between new Kow (EC-JRC-2018) and original USEtox® values (top: showing the different data source; bottom: showing the difference in log between values)
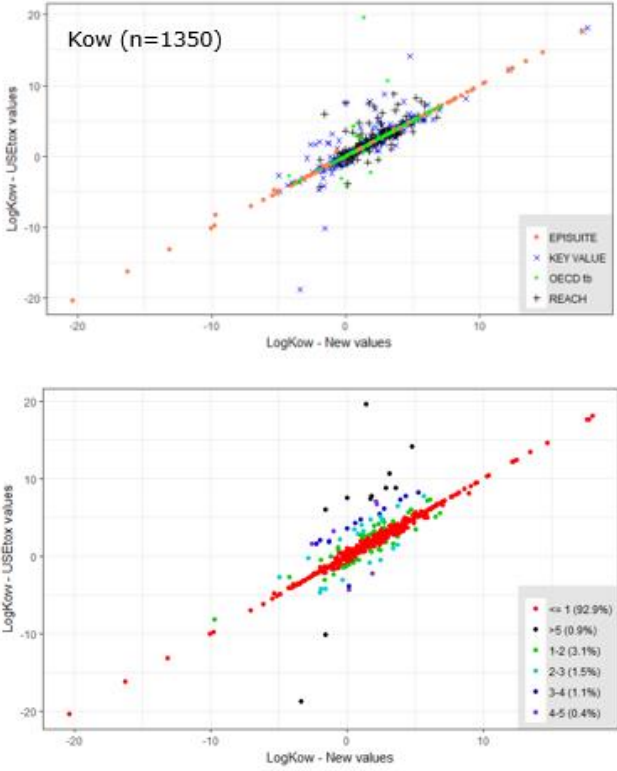

Figure 8: Comparison between new Vapour pressure (EC-JRC-2018) and original USEtox® values (top: showing the different data source; bottom: showing the difference in log between values)
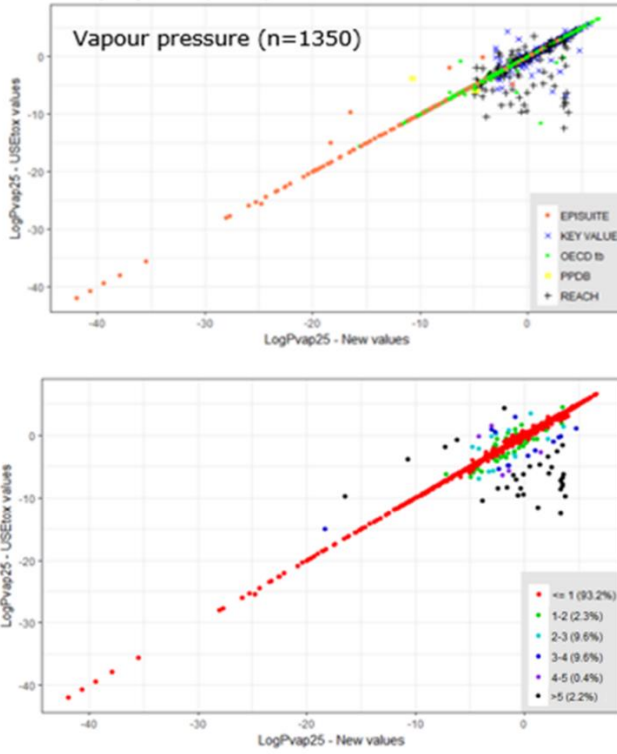
Figure 9: Comparison between new Henry's law constant (EC-JRC-2018) and original USEtox® values (top: showing the different data source; bottom: showing the difference in log between values)
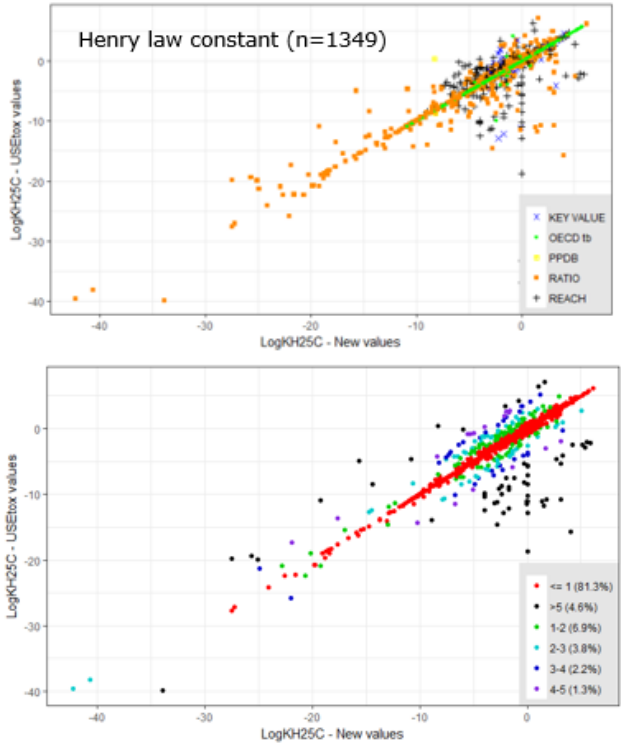
Figure 10: Comparison between new Koc (EC-JRC-2018) and original USEtox® values (top: showing the different data source; bottom: showing the difference in log between values)
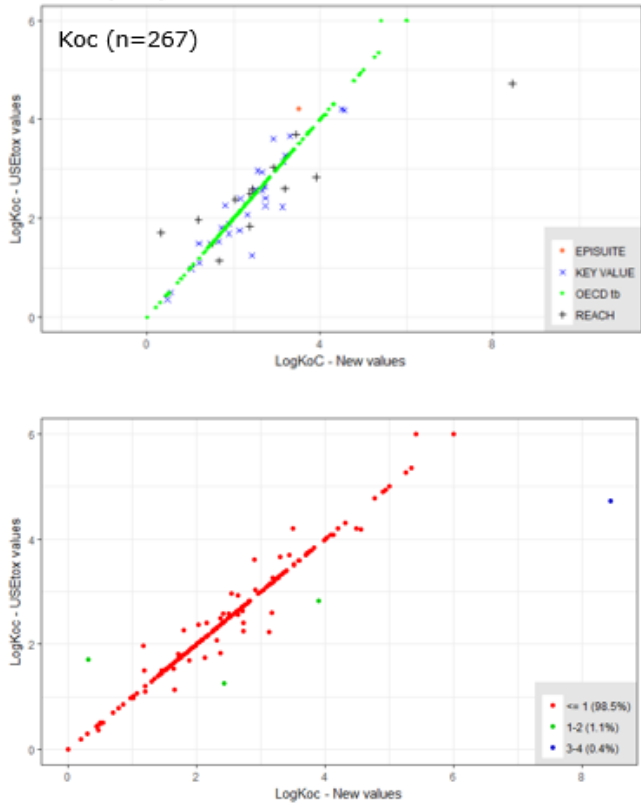
Figure 11: Comparison between new Water solubility (EC-JRC-2018) and original USEtox® values (top: showing the different data source; bottom: showing the difference in log between values)
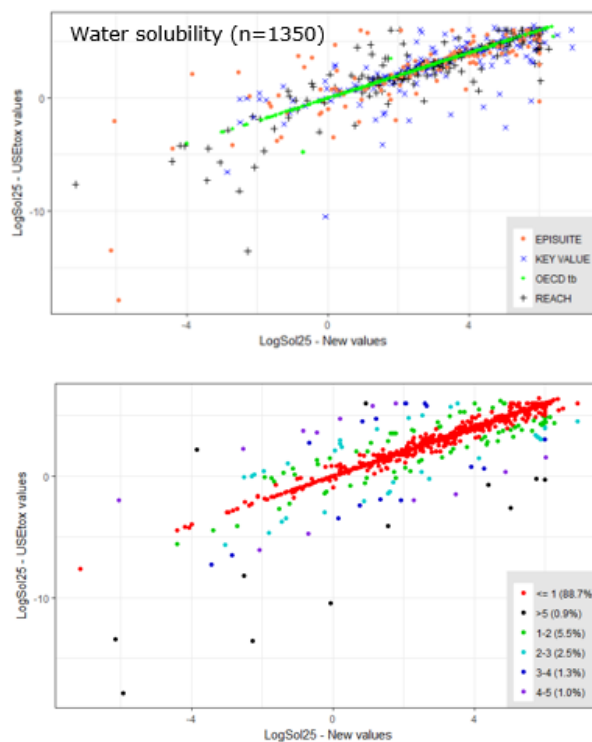
Figure 12: Comparison between new Water biodegradation (EC-JRC-2018) and original USEtox® values (top: showing the different data source; bottom: showing the difference in log between values)
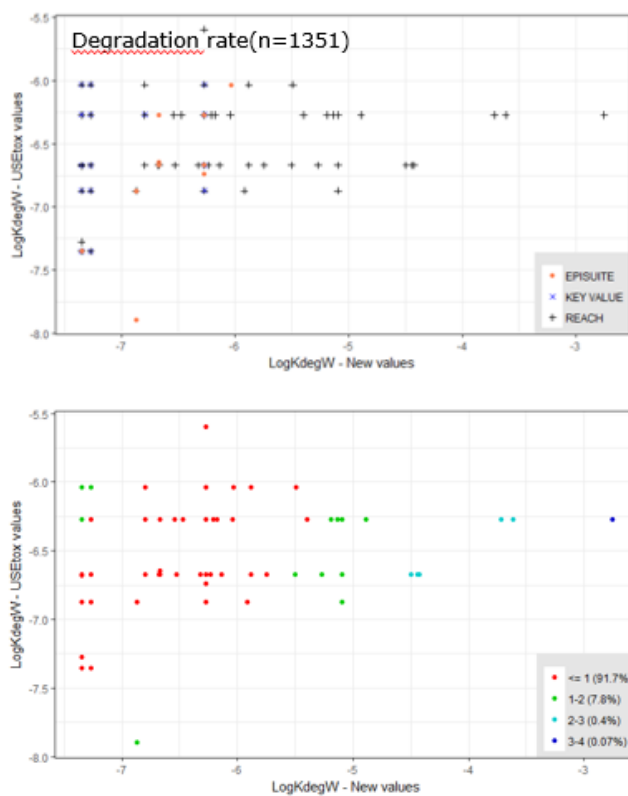
The following observations can be made:

— For Kow and Vapour pressure there is 1 to 1 agreement between the EPIsuite™ and the OECD toolbox prediction, however, a higher variability is visible for the Henry's law and water solubility prediction. This is probably due to the use of the same equation for the prediction of the first two parameters, while different predictions are used for the Henry's law constant and water solubility.

— Kow and Vapour pressure values extracted from REACH-IUCLID database show overall a good agreement with predicted values, with some values being order of magnitude away from the prediction (for both 'key value' extracted from CSA and values extracted using the procedure described in this report).

— For Koc, measured values are well aligned with predicted values.

— For water solubility, measured values do not correlate with predicted values suggesting that the use of predicted values as done in the USEtox® model may under or overestimate the 'true' solubility of the substances.

— For biodegradability in water, substances are distributed on few horizontal lines corresponding to the default rates assigned to each Biowin3 outcome categories (data used in USEtox®); while data on the x-axis are more widely distributed reflecting the experimental rates extracted from REACH-IUCLID database.

— The percentage of substances with a ratio lower than one order of magnitude is always higher than 80%, ranging from 81.3% (Henry's law constant) to 98.5% (Koc).

These observations suggest that using measured data over predicted will impact significantly the calculation of the fate and exposure factors via the USEtox® model. The two parameters showing the highest difference being the water solubility and biodegradation rate. Furthermore, the high amount of Koc experimental data available in REACH-IUCLID database represents a strong improvement produced with this work as these parameter was essentially estimated from Kow in USEtox®. Lastly, pKaChemClass predicted with ADMET® predictor was compared with the one in USEtox® (generated with SPARC® 6.0) for the 1350 substances in common (Table 16).

Table 16: Comparison between ADMET® and USEtox® pKaChemClass

ADMET® ⟶ pkaChemClass	Neutral	Acid	Base	Amphoter	Total
USEtox® pKaChemClass ↓					
Neutral	**594**	50	64	42	750
Acid	30	**218**	4	33	285
Base	2	5	**233**	16	256
Amphoter	0	5	2	**52**	59
Total	626	278	303	143	1350

Total agreement occurs for 1097 substances (81%), along the bisector of the matrix. The most concerning difference is the disagreement between acid and base class (9 substances). The majority of differences regard the "USEtox® neutral" class (156 classified as "neutral" are distributed in other classes) and the "ADMET® amphoter" class (91 substances classified as "amphoter" but with different category in USEtox®). The agreement for more than 80% of the data suggests an equivalence in using ADMET® predictor or SPARC® 6.0 to retrieve pKa information for USEtox® model.

4 Deriving substance aquatic toxicity hazard values from the REACH, EFSA and PPDB

The R code used to retrieve aquatic toxicity from the REACH-IUCLID, OpenFoodTox and PPDB databases and the list of variables available for each parameters are available in the online supplementary material (see annex 1 for list of supplementary material and http://eplca.jrc.ec.europa.eu/LCDN/developerEF.xhtml). However, the final individual species toxicity values are not available on the online materials for property and confidentiality reasons, but they are available on the ECHA dissemination website. Only the final substances hazard values are made available.

4.1 The REACH ecotoxicity data

All aquatic ecotoxicity data present in the REACH-IUCLID database as of May 2015 were exported by ECHA from the International Uniform Substance Information Database (IUCLID 5.5 (ECHA, 2018a)) into several Excel files.

The ecotoxicity database contained 305068 toxicity results on 7714 substances. The database includes data from acute and chronic toxicity tests performed with various taxonomic groups, derived with or without using regulatory-adopted testing guidelines, derived from QSAR (Quantitative structure activity relationship) and read-across methods, or obtained from the scientific literature.

The extract includes all substances registered from the first two of the three official registration deadlines (2010, 2013 and 2018) for substances already in use at the time of the REACH enforcement. Registrations for the last deadline (June 2018) covers low tonnage substances for which limited test data are required. Therefore, new ecotoxicity data available onwards from June 2018 are expected to be rather limited in number and relative relevance. Nevertheless, the procedure we are proposing herewith, via R programming applied to the excel files that the data were downloaded into, can be reapplied automatically to new extracts of the REACH-IUCLID to take advantage of new data or existing dossier updates.

Each row of the ECHA-exported Excel files is dedicated to the characteristics and results of a single toxicity test. Each of the 28 columns provides experimental details such as duration, reliability codes, adequacy codes, type of study, guidelines, etc. and corresponds to a specific data field in IUCLID 5.5. The number of available data fields in IUCLID 5.5 for each test is much higher than 28, but only the entries that were judged to be important to understand the context, quality, and results of the toxicity study for the present purposes were retained (Table 17).

Table 17: List of information provided with each test results

1	DOSSIER_UUID number
2	Study Report number
3	Duration
4	Duration Unit (in sec, min, hours, weeks, etc.)
5	Reference point (EC50, NOEC, LOEC, etc..)
6	Additional information regarding the endpoint: based on biomass or growth, temperature, hardness, specific test conditions).
7	Low range test results qualifier: 0, >, >=, empty, blank
8	Low range test result
9	High range test results qualifier: 0, <, <=, empty, blank
10	High range test result
11	Unit of the reported results (mg/L, kg/L, v/v, etc., in total several dozens of reported units.

12	Additional information on unit
13	Indication if the test results were based on nominal or measured concentration
14	Information on test material: active ingredient, dissolved, element, labile, etc.
15	Additional information on units
16	Reliability (cf. Klimisch score (Klimisch et al., 1997))
17	Species name
18	Additional info on species
19	Study Type: experimental, QSAR, read across
20	Additional info on study type
21	Adequacy: Key study, Supportive study, weight of evidence, etc.
22	Salinity
23	pH
24	Water Media: freshwater, salt, marine, etc.
25	Guideline
26	Additional info on guidelines
27	EC Number
28	CAS Number

The export of the original REACH-IUCLID database presented several practical challenges, such as information mistakenly recorded in the wrong column, missing test duration or test results, spelling mistakes in species names, lack of phylogenetic information (family, class, phylum, etc.), and durations and test results expressed in different units, etc.

The first data curation operation consisted of deleting duplicate records (42613) and records where duration (4418), results (8658) and species names were absent (11068), bringing the database to 242729 toxicity records from 94199 different study reports. The list of corrections applied to the original database as well as the corresponding R codes are provided with the online version of the report.

Approximately 65% of the substances were registered as "mono-constituent", 26% as "Unknown or Variable composition, Complex reaction products or Biological materials" (UVCB), and 9% as "multi-constituents" (Table 18). Organic and inorganic substances dominate the database with a few compounds in the "petroleum", "organometallic", and "element" categories. However, both the 'composition' and the 'type of substance' are information entered by the REACH registrants and are not always accurate. For the same substances, some registrants described the substance as mono-constituent, while others as multi-constituent or UVCB (marked as 'tie' in Table 18). The same observation applies to the definition of organic, inorganic, element, organometallic, etc. We have applied the 'majority voting rule' (i.e., that the nature of the material is based on the component present in the greatest amount) to propose one single descriptor for each substance. In case of doubt, the information added from the OECD toolbox was used to assign a substance composition (OECD, 2018). These entries do not change the toxicity information, but only affect processes like sub-grouping of compounds, with associated changes in efforts to describe patterns in the database.

Table 18: Total number and type of substances REACH-IUCLID (7713 substances).

Composition	Type of substances	Number of Substances	%
mono	element	21	0.27%
	inorganic	603	7.82%
	no info[1]	153	1.98%
	organic	4095	53.09%
	organometallic	103	1.34%
	petroleum product	28	0.36%
	tie	22	0.29%
multi	inorganic	39	0.51%
	no info	26	0.34%
	organic	593	7.69%
	organometallic	9	0.12%
	petroleum product	3	0.04%
no info	organic	1	0.01%
tie	inorganic	4	0.05%
	no info	1	0.01%
	organic	40	0.52%
	organometallic	1	0.01%
	tie	4	0.05%
UVCB[2]	inorganic	116	1.50%
	no info	104	1.35%
	organic	1321	17.13%
	organometallic	46	0.60%
	petroleum product	372	4.82%
	tie	8	0.10%

1 Substance origin not available in the REACH-IUCLID database;
2 UVCB (Substance of Unknown or Variable composition, Complex reaction products, or Biological materials)

Although the total number of different biological species available in the dataset was 993, the most tested taxonomic groups are: crustaceans, fish, and algae (88% of the results and study report), related to preferred standard tests (Table 19). Approximately 78% of the substances have at least one toxicity value for each of these three taxonomic groups.

The species group called 'others' is composed mainly of bacteria and less common test species. The taxonomic groups listed in table 19 (left column) were taken from the list of suggested groups to be included in Species Sensitivity Distribution models standardized for use in substance safety assessment (EC-JRC, 2003).

For the majority of substances, toxicity data are available for at least 3 species (usually from fish, crustaceans and algae), but for about 1500 substances (mainly organic substances) only 1 or 2 toxicity test data are available (Figure 13). A detailed investigation would be needed to explain why many substances have less than the usual substance safety assessment requirement of minimal three toxicity data. The distribution in figure 13 has implications for the derivation of the Species Sensitivity Distribution (EC-JRC, 2003; Posthuma et al., 2002) underlying the derivation of the effect factors as the number of species from different taxonomic group is recommended to be at least higher than 8 according to EC-JRC 2003.

Table 19: Phylogenetic composition of the REACH-IUCLID database and number of substances, study reports, and test results per taxonomic group.

Taxomonic groups	Phylum	Order	Family	Species	Substances	Study reports	Test results
Crustaceans	1	15	58	183	7387	32594	78654
Fish	1	22	62	212	6685	29886	75421
Algae	12	51	74	180	6894	20319	59667
Amphibians	1	2	7	34	189	588	2040
Anellids	1	12	19	30	248	1119	4106
Insects	1	7	25	52	329	1062	2624
Molluscs	1	19	35	97	447	3008	6725
Others	19	61	87	159	1244	4286	8129
Plants	4	14	17	36	399	1020	3350
Rotiferans	1	2	5	10	190	317	917
Total	**42**	**205**	**389**	**993**	**7713**	**94199**	**241633**

For most taxonomic groups, three species represent more than 60% of the available toxicity tests for that taxon (up to 83% for crustaceans) (Table 20). The complete list of species per taxonomic group is available in the online version of the report.

Figure 13: Number of substances tested plotted against the number of taxonomic groups tested per compound, discriminating the selected types of substances available in the REACH-IUCLID ecotoxicology database.
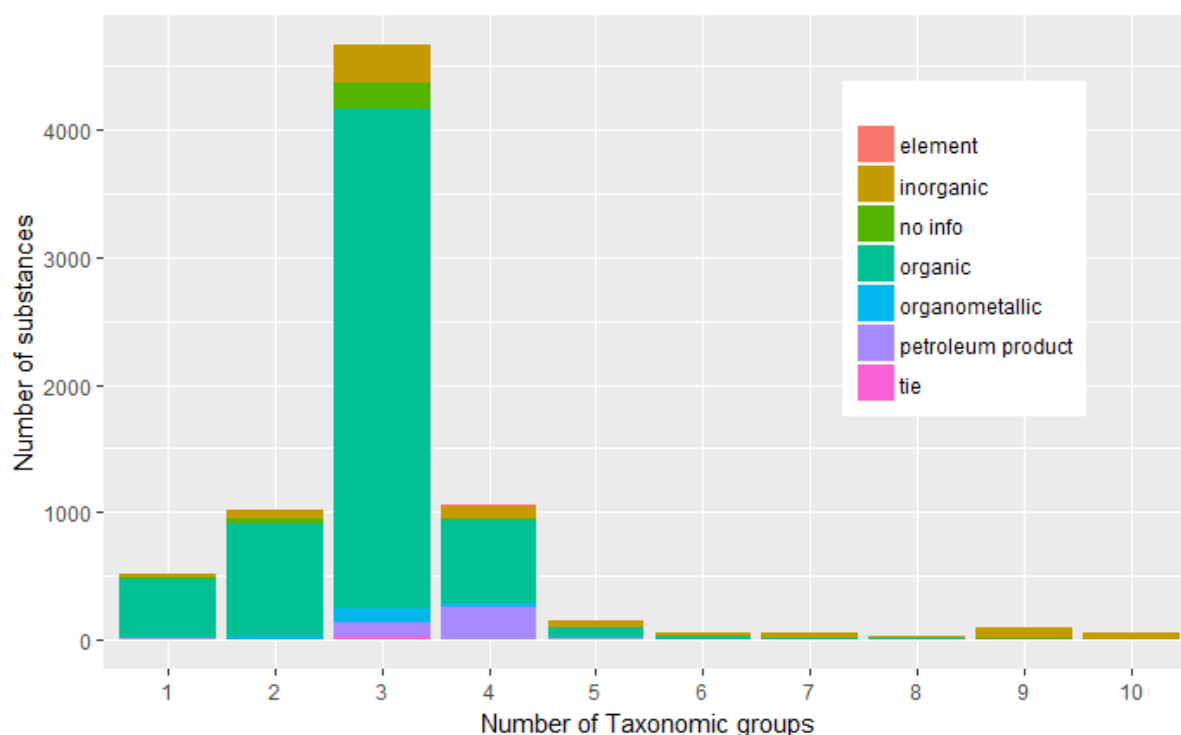
For the majority of study reports (65%), 1 to 4 toxicity test results were recorded, usually because of reporting multiple reference points (e.g., NOEC, EC10, EC50, etc.) or durations (e.g., 48h, 72h, 96 hr) per test. However, a significant number of study reports contained up to 20 reported toxicity values (from the same experiment). This is usually due to:

— More than 1 replicate per test (the same experimental conditions have been tested more than once).

— Different experimental conditions due to variations in: water hardness, pH, dissolved organic carbon (DOC), temperature, light intensity, etc.,

— Toxicity values based on nominal and measured concentrations, or based on the dissolved fraction or total concentration of the substance tested (usually for metal and inorganic).

Table 20: Three first dominant tested species per taxonomic group on the freshwater ecotoxicity database.

Taxonomic groups	Three most dominants tested species	% of test results per taxon
Algae	*Raphidocelis subcapitata*	50%
	Desmodesmus subspicatus	26%
	Skeletonema costatum	4%
Crustacean	*Daphnia magna*	66%
	Ceriodaphnia dubia	15%
	Americamysis bahia	2%
Fish	*Oncorhynchus mykiss*	26%
	Pimephales promelas	22%
	Danio rerio	19%
Amphibian	*Xenopus laevis*	38%
	Gastrophryne carolinensis	16%
	Anaxyrus terrestris	11%
Anellidae	*Tubifex tubifex*	47%
	Neanthes arenaceodentata	20%
	Aeolosoma sp.	7%
Insect	*Chironomus tentans*	36%
	Chironomus riparius	20%
	Chironomus dilutus	5%
Mollusca	*Magallana gigas*	17%
	Lymnaea stagnalis	12%
	Mytilus galloprovincialis	11%
Others	*Pseudomonas putida*	18%
	Strongylocentrotus purpuratus	10%
	Dendraster excentricus	10%
Plant	*Lemna minor*	60%
	Lemna gibba	23%
	Spirodela polyrrhiza	6%
Rotifera	*Brachionus calyciflorus*	65%
	Philodina acuticornis	11%
	Philodina rapida	11%

4.1.1 Curation of the data for the EU Environmental Footprint

Not all the data available in the REACH-IUCLID database are suitable for the purpose of deriving substance hazard values to be used with the USEtox® model and the EU Environmental Footprint.

Depending on the goal of the work, selection rules can be applied to select appropriate endpoints and tests for the specific assessment target. The information displayed in Table 17 was used to extract the endpoints considered valid to calculate final substance effect values for use in the EU-EF approach. The following sections describes the rules used (see online material for the R codes).

Rule 1: Selecting high quality data and type of test

To facilitate the use of the data, REACH requests that each record should be assessed for it adequacy for risk assessment, classification and labelling, etc. and for its reliability (inherent quality) using the Klimisch scoring system (ECHA, 2014; Klimisch et al., 1997).

For 'adequacy', test records are classified either as 'key study' (46% of the REACH-IUCLID database), 'Supporting study' (30%), Weight of evidence (14%), or 'disregarded study' (3%). Seven percent of the test records were not classified for 'adequacy.

For 'reliability' four levels are used to discriminate quality: k1 for 'reliable without restriction' (36%), k2 for 'reliable with restriction' (50%), k3 for 'not reliable' (8%) and k4 for 'not assignable (2%). Three percent of the test records had no Klimisch score.

All toxicity tests described either as a 'Key study', 'supporting study' and 'weight of evidence' were retained. The test results without any 'adequacy' descriptions were also retained if they are ranked 'k1' or 'k2' according to the Klimisch scores. 55% percent of the registered toxicity results are from experimental studies, while 39% are from read-across methods and 2% from QSAR approaches. Five percent have not been documented. All these study types have been retained as long as the record was classified as Klimisch scores k1 or k2.

Rule 2: Selecting freshwater media

The REACH-IUCLID database covers test performed in freshwater (80%), saltwater (11%), and brackish water (2%) media, with 9% of the data (21,569 test results) having no information on the exposure medium. Since the REACH regulation does not require test to be performed in saltwater, when exposure media information is missing, it was assumed that these tests were performed in freshwater (default situation). Since the purpose of our work is to provide substance toxicity values for freshwater ecosystems, only test in or assigned to freshwater media are retained.

Rule 3: Test values presented as ranges

Although for the majority of test records, only one value was registered, for a significant number of records, test results are displayed in 2 columns with low and high value ranges without any such qualifier or with qualifier such as: =, >, <, >=, <=, ca. (approximately) (Table 21).

Table 21: Number of test values recorded with or without qualifiers for the test reference point (effect concentration) in the REACH-IUCLID database.

Lower values		Higher values	
Qualifier	*Number of results*	*Qualifier*	*Number of results*
>	39602	<	4397
>=	8068	<=	1406
ca.	3493	ca.	59
=	190470		
Total	241633		5862

A large majority of the results have a numeric value in the low range with a qualifier =, ca., >=, or >. In contrast, only a few tests have their results expressed in the higher ranges (5862 test results). The following selections were made to maximize the use of available data:

— When there is a lower range value with the descriptors '>=, ca., or empty', the lowest value is selected. If, within this group, a test has also a higher value, this higher value is ignored.

— All lower range values described as '>' are ignored (n = 39602), unless the higher value is described as '=<' (n= 80 observations). In case of NOEC > than, the value was kept since it is still representing a concentration with no observed effect.

— All higher values described as '< than' are ignored, unless the lower range value is described as '>='. Then the lower value is used.

— When a lower range value is missing (0 or blank) and a higher value is available described as '<=', the higher value is used.

— When a lower value is described as >= and the higher value is described as <=, the lowest value is used.

— Values expressed as '<' are excluded (4397 test results).

Rule 4: Selecting Acute and Chronic effect values

To calculate a unique effect factor for each of the substances to be used in the Environmental Footprint approach, a clear separation between acute and chronic toxicity data is required to allow derivation of toxicity estimates based on acute or on chronic data only, or on both (which may be realized via application of acute to chronic extrapolation factors). In the IUCLID database, two different sections are used to report acute (short term) and chronic (long term) toxicity results for fish and aquatic invertebrate. In contrast, for algae, plants and other aquatic organisms only one section is used.

There are four aspects to determine if an individual toxicity test is an acute or chronic study: biological effect, reference point, duration and species. These four attributes should be in principle combined for every study to assign the test data to the acute or the chronic group. It is clear that assignment to the acute or chronic groups can be difficult for some test data based solely on the data in the IUCLD database. However, this assignment is required to maximize the use of each toxicity test (i.e., Environmental Footprint process).

Biological effect

Biological effects are usually recorded in the IUCLID database in the result sections and may occur in multiple different subsections. Except for algae tests where effects based on 'biomass' or growth' were partially recorded, for the large majority of toxicity test results, this type of information was not available in the REACH extract obtained from ECHA.

Reference points

Ecotoxicity tests were reported using 59 different reference points with the most frequently used being LC50, NOEC, EC50, EC10, LOEC, etc. (Table 22 and online material).

Table 22: Toxicity reference points and frequency of occurrence in the REACH-IUCLID database covering 99% of the reported results.

Reference point	Number of results	%	Reference point	Number of results	%
LC50	54972	22.8%	EC90	854	0.4%
NOEC	51951	21.5%	TTC	592	0.2%
EC50	45621	18.9%	EL10	534	0.2%
EC10	17302	7.2%	LL0	514	0.2%
LOEC	14254	5.9%	ErL50	475	0.2%

Reference point	Number of results	%	Reference point	Number of results	%
EL50	9738	4.0%	EL0	442	0.2%
NOELR	5931	2.5%	EbL50	437	0.2%
LL50	5726	2.4%	ErC50	413	0.2%
EMPTY*	4368	1.8%	LC20	410	0.2%
LC100	4205	1.7%	LOELR	387	0.2%
LC0	3895	1.6%	IC10	384	0.2%
EC20	3439	1.4%	IC25	348	0.1%
EC100	3238	1.3%	EL100	338	0.1%
EC0	3,185	1.3%	EbC50	320	0.1%
NOEL	1,967	0.8%	ChV	299	0.1%
LC10	1,589	0.7%	MATC	258	0.1%
IC50	1,262	0.5%	IC20	223	0.1%

*Test results not associated with one of the reference points listed in the table (cell was empty).

LC: lethal concentration, NOEC: no observed effect concentration, EC. Effect concentration, LOEC: lowest observed concentration, suffix 'r' and 'b' stand for 'growth rate' and 'biomass', LL: loading rate, EL: loading effect, IC: immobilization concentration, ChV: chronic value, MATC: maximum acceptable toxic concentration, TTC: toxicity threshold of concern.

When more than one reference points was reported for the same test, the following rules were used:

— For acute and chronic median effect (50% effect) tests, reference points usually reported are EC50 (effect concentration), IC50 (immobilization concentration) and LC50 (lethal concentration). If a single test reports all three reference points, the order of preference was EC50 > LC50 > IC50. If, for the same substance / same species, one test reports an EC50 and a second test reports an IC50 (or LC50), both reference points were included in the calculation of a species geometric mean test value. For algae, the 50% effect can be based on growth rate (ECr50) or biomass (ECb50). If both values are reported, growth rate was selected.

— For chronic tests, reference points are more diverse and priority was given as follows: ECr10 > EbC10> EC10 to EC20 > NOEC > LOEC > MATC - ChV > TTC. If, for the same substance / same species, one test reports an LC10 and the second test reports NOEC (or any other chronic endpoint), both reference points were used to calculate a species geometric mean test value.

Duration

Within each taxonomic group, the reported duration of exposure varies from minutes to months for tests registered in the 'short-term' and 'long-term' IUCLID section (Figure 14, top). In the IUCLID database, users are invited to register their data either in the short-term or in the long-term section. Data inspections, however, made clear that there are inconsistencies within the database in decisions made as to whether individual toxicity tests were short (i.e., acute) or long-term (chronic) in duration. To provide consistency, IUCLID designations of short and long-term were ignored and rules were established to assign tests based on duration, reference point, biological effect and species.

For fish, algae and crustaceans, the most frequent duration corresponds to 96 hours for Fish Acute, 28 days for Fish Chronic; 48 hours for Crustacean Acute, 21 days for Crustacean Chronic; 72 hours Algae for acute (as EC50) and chronic (as NOEC or ECx)(Figure 14, bottom).

For the algae, it could be argued that the EC50 determined at 72 hours is a chronic endpoint (algal cells divide many times in 72 hours), although - from a regulatory point of view - they are considered acute and NOEC or EC10 are considered chronic endpoints. The use of

duration limits to separate acute from chronic is consistent with a recent attempt to use REACH data to calculate USEtox® substance hazard values (Müller et al., 2017).

The standard recommended test durations are based on OECD (Organisation for Economic Co-operation and Development), ASTM (American Society for testing chemicals), EPA (US Environmental Protection Agency) and other standard aquatic toxicity test methods which use specific durations to assign acute and chronic exposures.

Figure 14: Available Duration of toxicity test exposures in hours for each taxonomic group. The violin plot shows the full distribution of the test duration data for the different 'groups'. The size of diamond shapes for each species indicate the relative number of data reported for a specific duration. The vertical lines represent the range of the data.
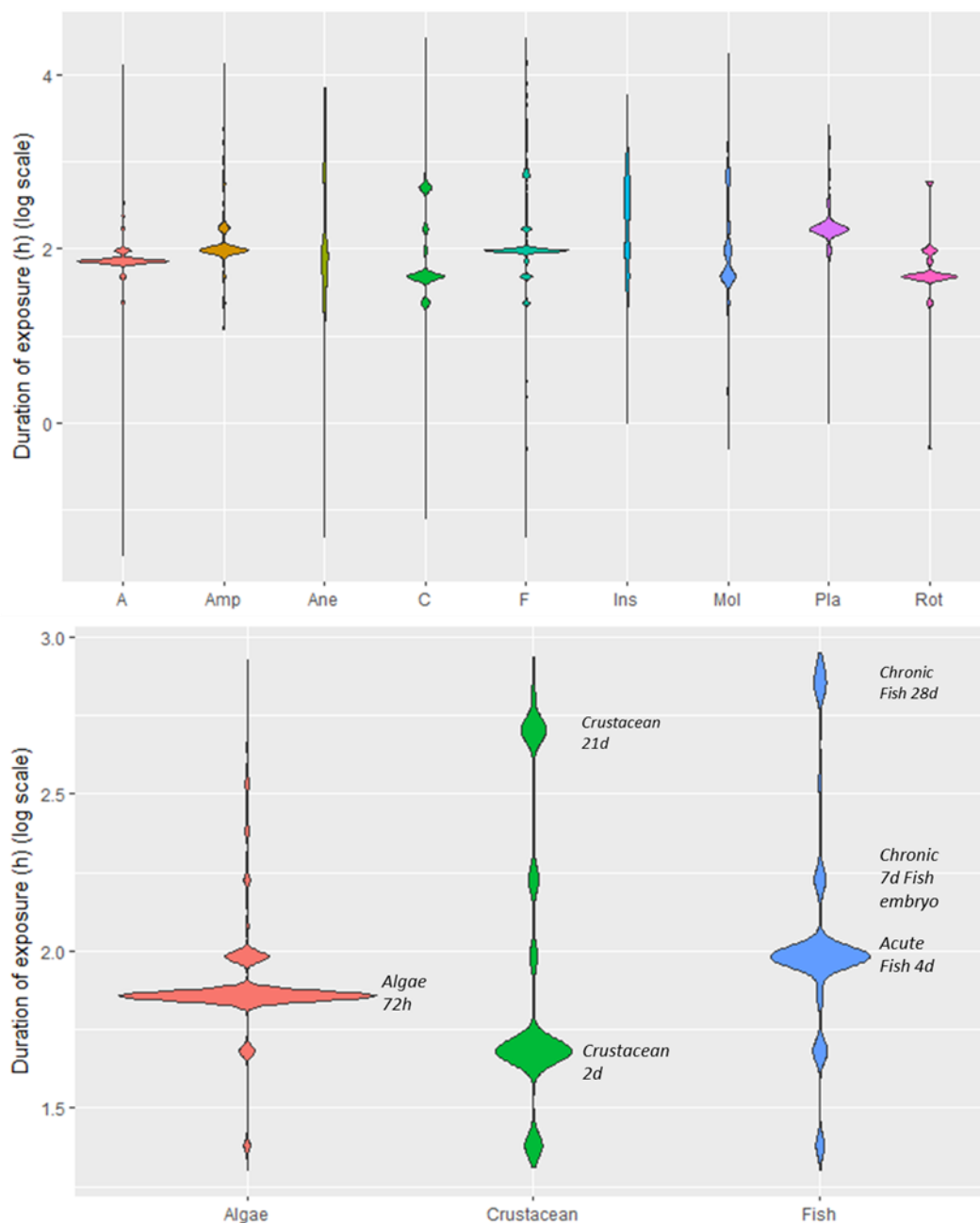
Top graph, A: algae, Amp: amphibians, Ane: anellids, C: crustaceans, F: Fish, Ins: insects, Mol: molluscs, Pla: plants, Rot: rotifers

For each taxonomic group, a specific range of exposure (together with endpoint and biological effect) was used to pool the toxicity data into acute and chronic exposure categories (Table 23). Reported exposure durations typically match the official standard recommended durations, allowing for some variation due to the need to stop or prolong a test due to test observations and practical concerns (e.g., staffing). The use of a duration range to select assign acute/chronic ensures that no data are excluded because the duration may vary from the standard. For fish, invertebrates, and algae, those values align with standard test guidelines. For the other taxonomic group, the selection of the acute and chronic durations was based on the 'split' observable on figure 14 and from expert judgment. Assignment of rotifer toxicity data into acute or chronic was based solely on endpoints.

Table 23: Proposed endpoint and duration ranges to distinguish between acute and chronic exposure per taxonomic groups

Taxonomy groups	Acute reference points and duration	Chronic reference points and duration
	reference points EC50eq: EC50, LC50, IC50	reference points Chronic EC50eq: EC50, LC50, IC50 reference points Chronic NOECeq: EC10 to EC25, LC5 to LC25, NOEC, LOEC, MATC, ttc, ChV
Algae	≥ 40h and ≤ 120h	
crustaceans (mainly Daphnia)	≥ 40h and ≤ 120h	≥ 168h
fish	≥ 40h and ≤ 120	≥ 168h
molluscs	≥ 24h and ≤ 96h	> 96h
amphibians	≥ 24h and ≤ 96h	> 96h
annelids	≥ 24h and ≤ 96h	> 96h
Insects	≥ 24h and ≤ 96h	> 96h
Plants	≥ 48h and < 120h	> 120h
Rotifers	≥ 40h	

Rule 5: Using 'Measured' or 'Nominal' test concentration

For the majority of the test results (44%), the toxic effect concentration was expressed as nominal (i.e. the target concentration at the start of the test) while 39% of the test results were reported as measured concentration (analytically verified test substance concentrations) (Table 24).

In the context of a substance safety assessment, it is critical to base the effect value on the most relevant tested concentration (i.e. measured) as some compounds can be (bio)degraded, volatilized or adsorbed to test vessels. When toxicity data based on measured concentrations are not available, results from nominal concentrations provides the next best data for use in the assessment.

Similarly, for the purpose of the EU-EF, where products are compared to each other, it is important to have data on as many substances with toxicity data as possible. Retaining measured concentration only, would eliminate 44% of the data in the database. Therefore, we pooled nominal and measured concentration. However, if in the same test toxicity values were reported both for nominal and measured concentrations, measured values were selected as priority. For some substances, such as metals, it was essential to base the toxicity assessment on the fraction of metals dissolved in the water media. In this case, if for the same test values were reported for both total and dissolved fractions, the latter was systematically retained.

Table 24: Number of test results expressed as a specific type of substance measurement in the test media for the REACH-IUCLID database

Concentration as:	Number of test results	%
nominal	106513	44%
measured (arithm. mean)	43250	18%
measured (not specified)	30908	13%
Empty	28514	12%
measured (geom. mean)	10652	4%
measured (initial)	8317	3%
no data	5568	2%
Estimated	4831	2%
measured (twa*)	2886	1%
acid equivalent	194	0%

*twa: time-weighted average concentrations

Rule 6: Dealing with test replicates and different test conditions.

The information regarding test replicates and test conditions (i.e., testing different temperature, different pH, different DOC, etc.,) are recorded in a 'free text' field (column n° 6 and 15 - Table 17) without a predefined structure that could allow automated selection and treatment of different experimental test conditions.

For the calculation of the substance toxicity value, we opted for arithmetic means of all tested conditions (average of replicates, average toxicity for all water hardness tested, average toxicity for all DOC tested, etc.) since different test conditions normally represent to some extent the diversity of situations in the real environment. Taking an average is a way to acknowledge that the diversity of real environmental conditions can be represented by the average data point.

4.1.2 The final ecotoxicity data selected from the REACH-IUCLID database

The selection procedure to build the final ecotoxicity database is summarized in figure 15. Using the rules described previously, two successive database versions were created from the initial REACH-IUCLID database.

The first one containing only high quality data performed on freshwater species (rules 1 to 3) with 154583 toxicity test records (about 50% of the initial REACH-IUCLID database). Since these modifications are coded in R, the selection procedure applied on the REACH-IUCLID can be easily modified to address different needs. For example, many test results were excluded because the reported value was expressed as 'higher than' or 'lower than' which is not useful information in a substance safety assessment, as these values cannot be used to precisely define a test metric and an associated equivocal substance safety conclusions. However, in the context of the EU-EF approach substance, these values may still provide interesting comparative information: if substance has a toxicity value > than 100 mg/L it could be concluded that it is not toxic to aquatic biota compared to those have a value of 0.1 mg/L.

Figure 15: Three steps applied to the initial REACH-IUCLID ecotoxicity database to build the final database from which substance-specific Effect Value for LCA could be derived
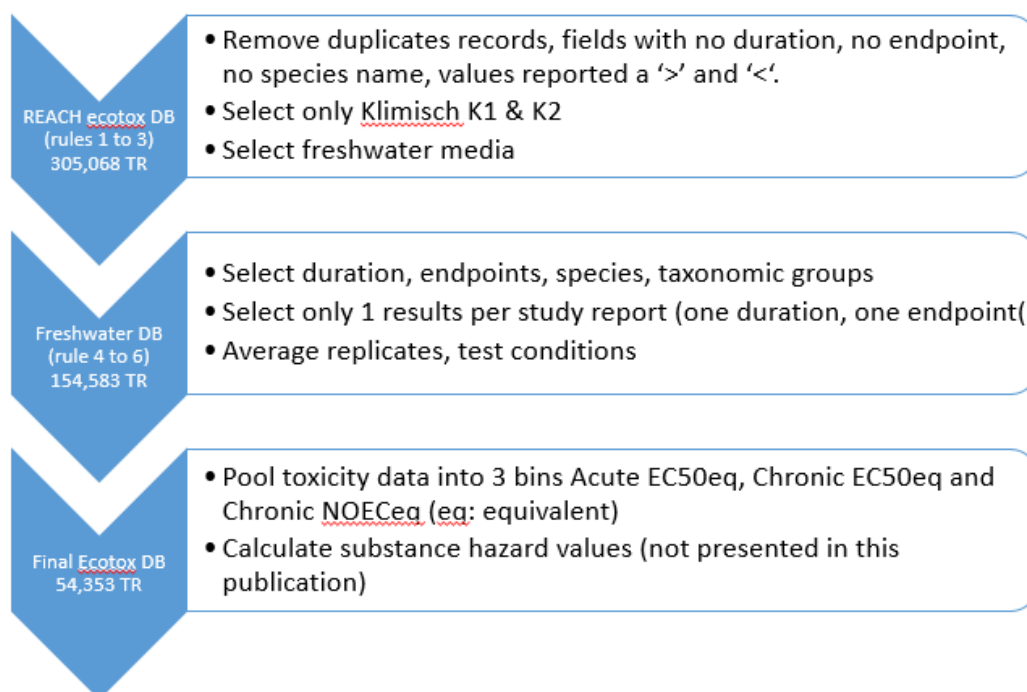
A second database (using rules from 4 to 6) contains the Acute and Chronic data that are considered of appropriate quality to be used to derive a hazard value for each of the substances. Toxicity data were then pooled into three categories: Acute EC50equivalent (29412 test results), Chronic EC50equivalent (3197 test results), and Chronic NOECequivalent (21744 test results) (Table 25 and Figure 16). This sub-set of selected data represents approximately 17% of the initial number of tests in the REACH-IUCLID database.

Table 25: Total number Acute EC50eq, Chronic EC50eq, and Chronic NOECeq toxicity in the final ecotoxicity database (total unique substance: 6461).

Taxonomic group	Acute EC50eq		Chronic EC50eq		Chronic NOECeq	
	Number of substances	Number of test results	Number of substances	Number of test results	Number of substances	Number of test results
Algae	3548	6528	na	na	4998	9772
Amphibians	52	213	33	148	43	209
Annelids	98	259	6	6	50	67
Crustaceans	3590	11098	886	1680	2468	5718
Fish	3153	10556	419	949	1366	4432
Insects	171	372	15	20	179	356
Mollusca	147	327	63	120	183	455
Plants	37	39	137	235	260	525
Rotifers	14	20	29	39	131	210
Total		29412		3197		21744

Figure 16: Endpoints reported in the EC-JRC-2018 ecotoxicity database based on the REACH data.
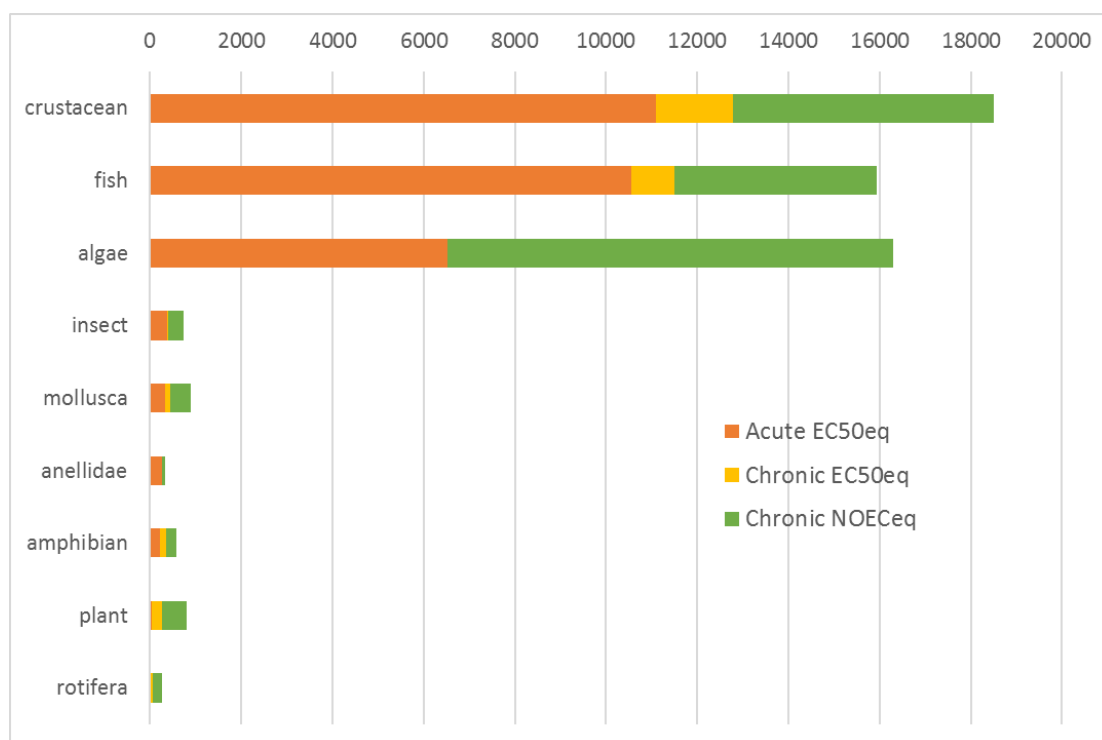
Muller et al. 2017 have also built a large ecotoxicity database aimed at comparing existing USEtox® hazard value (HC50) with the ones calculated using the REACH database. Although the selection procedure was similar to the one applied here, the study was restricted to those substances present in both the current USEtox® and REACH databases (i.e., 819 substances). In contrast, we applied the selection procedure to the whole REACH database (7713 substances) with the aim to eventually calculate new hazard value using the USEtox® approach (or any other LCA models) for as many substance as possible.

In contrast to present work, the work from Müller et al. 2017 used only EC50, IC50 and LC50 to build Acute and Chronic data bins, while we are proposing to extend the number of endpoints to make use of as much of the toxicity data generated and registered under REACH. Another important addition to the previous work is the creation of a new data bin using all existing chronic reference points such as NOEC, LOEC, ECx…. Those reference points represent 40% of the substance toxicity database (Table 25 and figure 16) and are considered to represent an appropriate reference point for deriving substance hazard value based on the preferred Chronic exposure data.

The approach used by USEtox® to retain only chronic EC50 (relatively rare data) and extrapolate acute EC50 to chronic EC50 via an extrapolation factor of 2 for all organic substances may have a lower reliability, due to the lower data numbers available, than the currently proposed approach for LCA (Müller et al., 2017).

For the three most tested taxa, the availability of reference points for acute and chronic exposure are presented in Figure 17. For algae, it could be argued that the EC50 determined at 72 hours is a chronic reference point (algae divide many times in 72 hours), although -from a regulatory point of view- they are considered acute and low effect levels like the NOEC or EC10 (also determined at 72 hours from the same test) are considered chronic reference points.

Figure 17: Main reference points reported for the fish, crustaceans and algae in the REACH ecotox database.
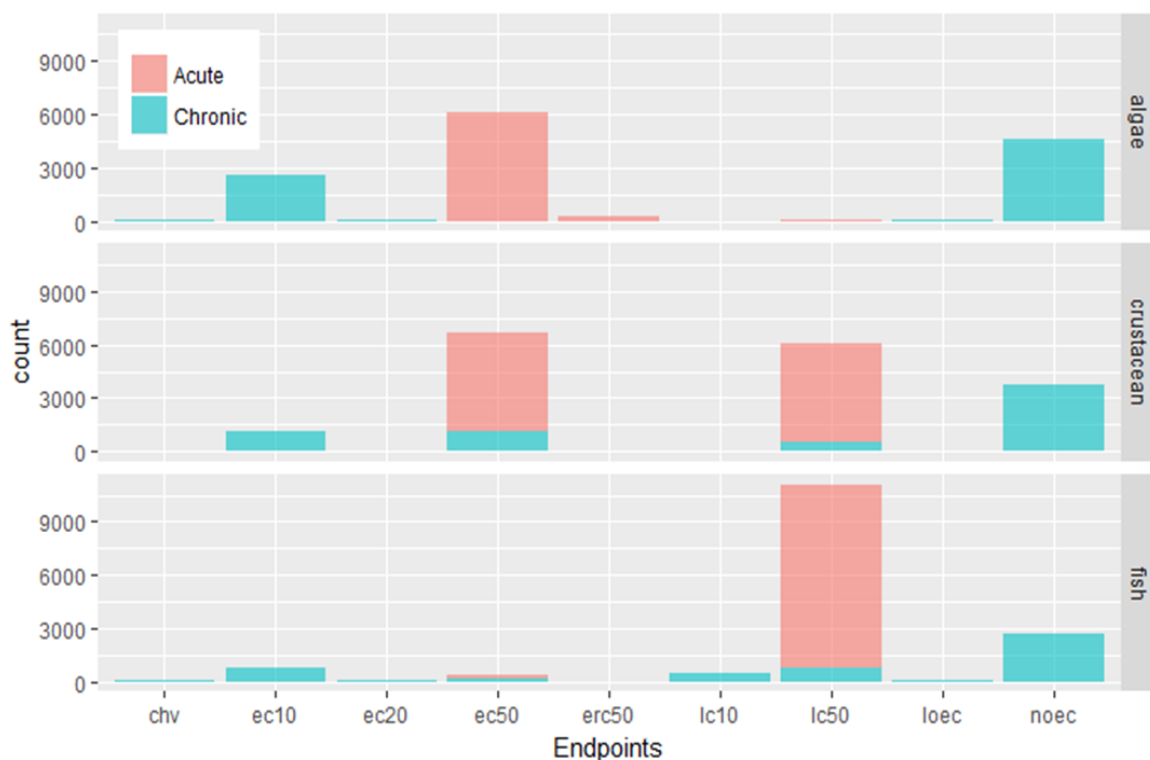
For Fish acute EC50eq, the dominant reference point in the final database is LC50 (96%) with the remaining tests reporting EC50 values. For crustaceans, the dominant reference point is EC50 (78%) with the remaining data points being LC50 (20%) and IC50 (2%). For algae, more reference points are available, mainly because EC50s can be reported based on biomass (ECb50) or growth rate (ECr50), but the dominant reference point is EC50 (96%).

The overlap between all retained reference points to describe acute toxicity suggests that building a substance-specific database aggregating slightly different reference points (regarding their name, not their interpretation) should be acceptable in the context of LCA / EU-PEF (figure 18).

For the Fish chronic NOECeq, 91% of the test results are expressed as a NOEC. Few test results are expressed as LOEC (2.5%), Chv (4.3%), or EC10 (1.4%). Surprisingly, the EC10 value appeared to be sometimes lower than the NOEC for compounds tested multiple times on a species (Figure 18). Further, the distribution of the LOECs overlap the NOECs.

For crustaceans, 88% of the data points are NOECs, while 9.4% are EC10 values. The rest of the reference points are EC16 and 20, LOEC and Chv.

For chronic Algae, 52% of the data points are reported as NOEC, and 45% as EC10 values. NOEC and EC10 overlap suggesting these two reference points as equivalent. With the comparisons of reference points within a taxa and exposure duration (acute, chronic), it is important to keep in mind that data points are obtained on a diversity of substances and species.

Figure 18: Distributions of selected toxicity reference points values for acute and chronic exposure for the main three taxonomic groups (fish, crustaceans and algae).
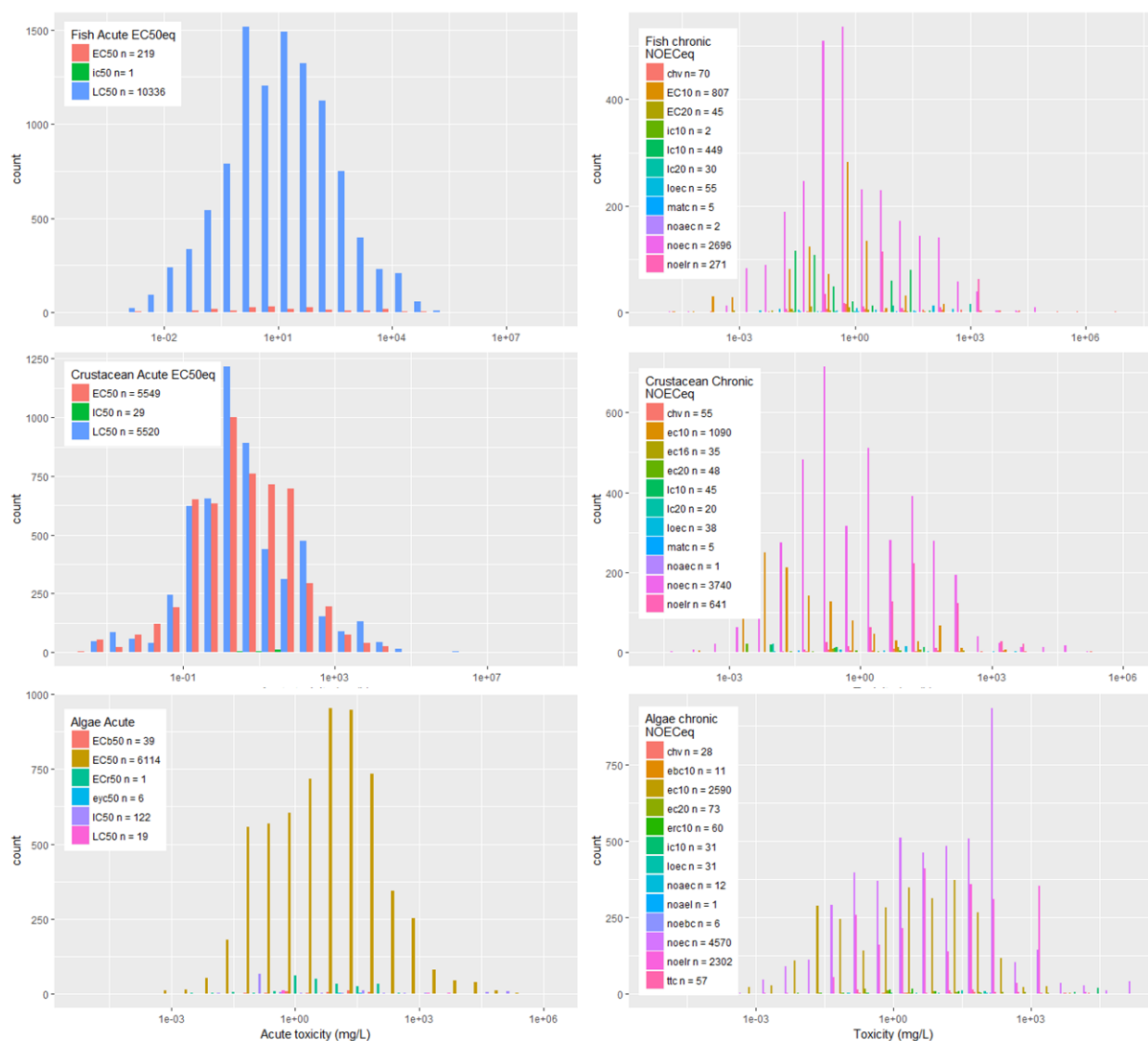
When performing a substance safety assessment, all the chronic reference points that were included in the Chronic NOECeq bins are commonly not considered equivalent. However, on the basis of numerical inspections it is becoming more and more evident that NOEC and EC10 to EC20 are estimates of the same type of sensitivity/effect information (Azimonti et al., 2015; Beasley et al., 2015).

The LOEC is usually not considered as equivalent to NOEC, since the LOEC corresponds to the next higher test concentration where a statistically significant effect occurs. However, in the substance current use, and based on the fact that LOEC are not so different from the rest of the reference points (see figure 18), we are proposing to include those reference points in the defined chronic NOECeq bin when no NOEC is available. The numerical impact of this choice on the Effect Values should be minimal as the number of LOEC is rather limited (9% of the chronic reference point).

An example of the data selection procedure is given in the online material for the substance formaldehyde (CAS: 50-00-0) to illustrate how data are finally selected.

In the initial REACH database, **97 toxicity results** were available for this substance.

After the first selection rules (Rule 1 to 3), **51 tests remained**:

— 21 test records were eliminated because no reference point was reported, although fourteen of those were classified as Klimisch k2 study.

— Two more tests were eliminated because the value was reported as 'greater than'.

— Thirty four test records were excluded because the tests were classified as Klimisch k3 and k4.

— Finally, nine tests were excluded due to the use of salt water as the test medium.

The selection of the final acute and chronic data resulted in a final database for the present purpose of **15 test results** (12 acute, 1 chronic EC50 and 2 chronic NOEC). Twelve of these test results are derived from one single value per test, while for three tests, the values is based on average on 2 or 3 replicates.

4.2 The 'OpenFoodTox' ecotoxicity database

Contrary to the aim of the REACH regulation, which is to collect and register into IUCLID all available data, the EFSA OpenFoodTox database take an opposite approach by recording in the database only values that are directly relevant for use in environmental and human risk assessment. Therefore, not all data available on plant protection products may be stored, but only the ones with high quality / relevance.

Limited intervention of the original data was required to assure its compatibility with the data extracted from the REACH-IUCLID database like unit conversion, exclusion of values reported as > or < than, exclusion of value generated from mixtures of formulae, and correction of few species names (to allow possible grouping with the REACH-IUCLID database).

From initially 2695 test results, and after the modifications described above, the database contains now 1956 individuals test results (1058 are chronic tests, 898 are acute) from which 33 tests are for salt / marine species. Those tests are, therefore, excluded for the calculation of the final substance Effect Factor (EfF).

All the selection and calculations have been made via the R program. All the codes used to extract the data starting from the OpenFoodTox database (Excel version) up to the final calculation of the Effect Factors, as well as graphs and tables will be made available.

The OpenFoodTox dataset contains:

— 408 unique substances with 2017 observations

— 578 fish tests

— 566 crustacean tests

— 364 algae tests

— 228 plants tests

— 199 Insect tests,

— 20 Mollusca tests

— 1 Annelidae test.

The majority of the 408 substances have at least 3 toxicity tests (Figure 19), while less than 50 have more than 4 tests available. Similarly, the majority of substances has been tested on 3 different species, while few have more than 5 species (Figure 20).

Table 26 and figure 21 show that none of the substances has the minimum number of SSD group to draw a reliable SSD curve. The large majority has 3 groups (137 substances have data Fish, Daphnia and Algae) while a significant portion has only 2 or even only 1 group of organisms tested. Substances for which the effect factor was calculated with at least 3 or more SSD group were considered of high quality, with 2 groups medium quality and with 1 group low quality.

Figure 19: Number of substance per number of ecotoxicological tests.
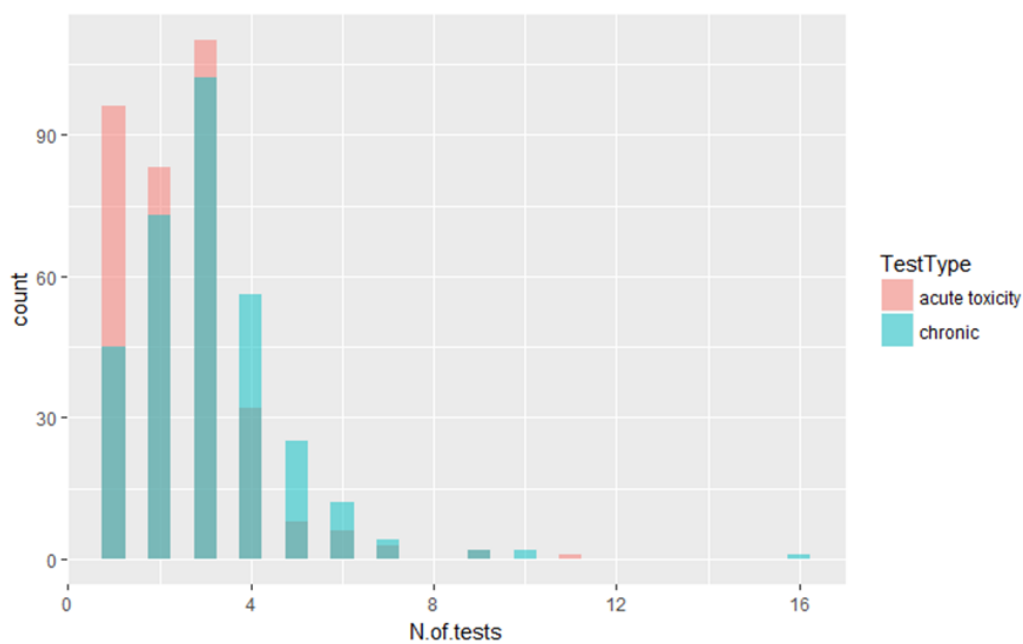
Figure 20: Number of substances per number of species tested.
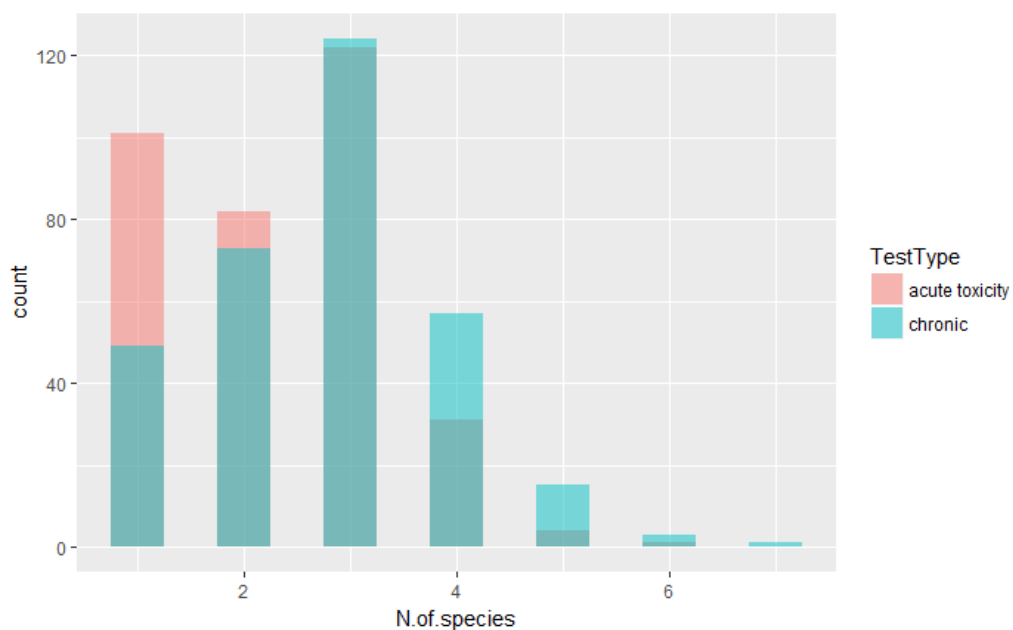
Table 26: Number of substances with number of available SSD group for acute and chronic data.

Number of SSD group available	Acute	Chronic
1 group	107	51
2 groups	95	73
3 groups	133	134
4 groups	6	56
5 groups		8
Total	341	322

Figure 21: Number of substances per number of SSD.group available.
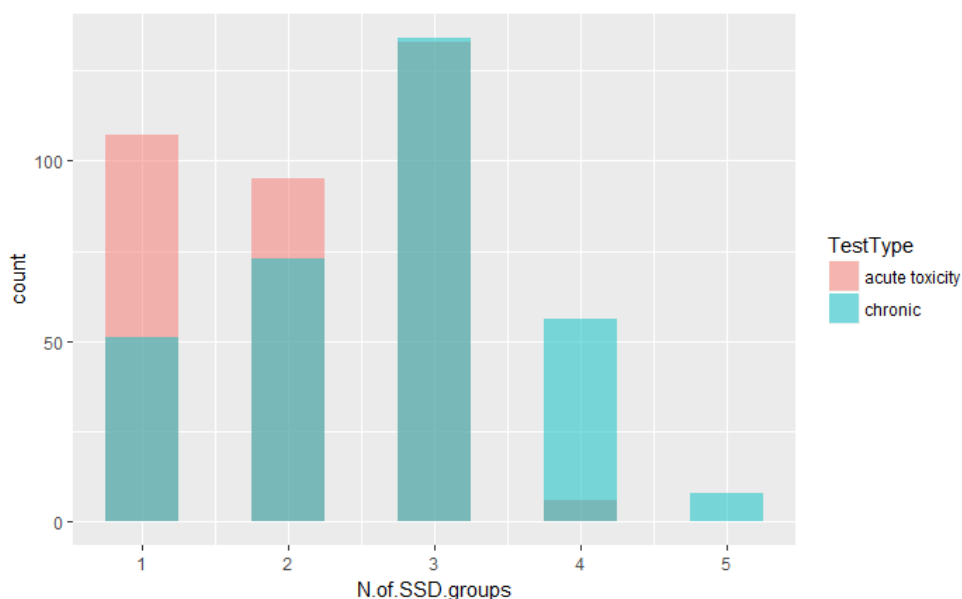
After calculating toxicity species geometric means, in case that for the same substance and same type of test (acute or chronic), several values were available for the same species, the EFSA data were pooled with the REACH-IUCLID database before the final hazard values were calculated (see chapter 4).

If a specific substance was both available in the REACH-IUCLID and OpenFoodTox database, with ecotoxicity data available on the same species, priority was given to EFSA data over REACH data. The PPBD database was only used to complete the EC-JRC ecotoxicity database for substances that were neither in REACH nor in EFSA.

4.3 The Plant protection ecotoxicity database

The PPDB contains pesticides data from various origins, including EU regulatory and evaluation data. The database was used only to complement the other sources to avoid redundant information.

The relevant ecotoxicological information contained in PPDB is reported in table 27. For each species, the chronic NOEC values were always prioritised when both chronic and acute values were available. Overall, information for 1316 pesticides are available in PPDB. Unfortunately, for the large majority of chemicals few species were tested. Therefore, for only few chemicals it was possible to draw a realistic SSD curve represented by many trophic levels (SSD groups).

Table 27: Relevant ecotoxicological information contained in PPDB

Species type (as in PPDB)	Reference point	Number of chemicals
Fish	Acute LC50 (96 hours)	1067
Fish	Chronic NOEC (21 days)	434
Invertebrates	Acute EC50 (48 hours)	1010
Invertebrates	Chronic NOEC (21 days)	507
Crustacean	Acute LC50 (96 hours)	218
Sediment species	Acute LC50 (96 hours)	74
Sediment species	Chronic NOEC (28 days)	200
Aquatic plants	Acute EC50 (7 days)	372
Algae	Acute EC50 (72 hours)	749
Algae	Chronic NOEC (96 hours)	72

4.4 Calculating substances aquatic toxicity hazard values

The UNEP-SETAC Pellston workshop in June 2018 (UNEP-SETAC, 2018) concluded with clear recommendation on how to derive substance hazard value to be used in the USEtox® model (Figure 22):

'*It is recommended to base effect modelling on a concentration domain of the SSD curve that is close to the domain of environmental (ambient) concentrations. Therefore it is recommended to use HC20 based on an SSD of chronic EC10-equivalents to estimate the potentially affected fraction of species (PAF).*

'*The chronic EC10-equivalent comprehends the chronic endpoints NOEC, LOEC, NOAEL, MATC, EC50, and chronic ECx where x is between 1 and 20, adjusted by appropriate correction factors. Specification of these is pending based on existing sources of literature. Acute to chronic extrapolations are used to fill in data gaps to increase coverage of species and substances.*'

Figure 22: Procedure for calculating effect factor in a given environment – UNEP-SETAC June 2018 Pellston workshop.
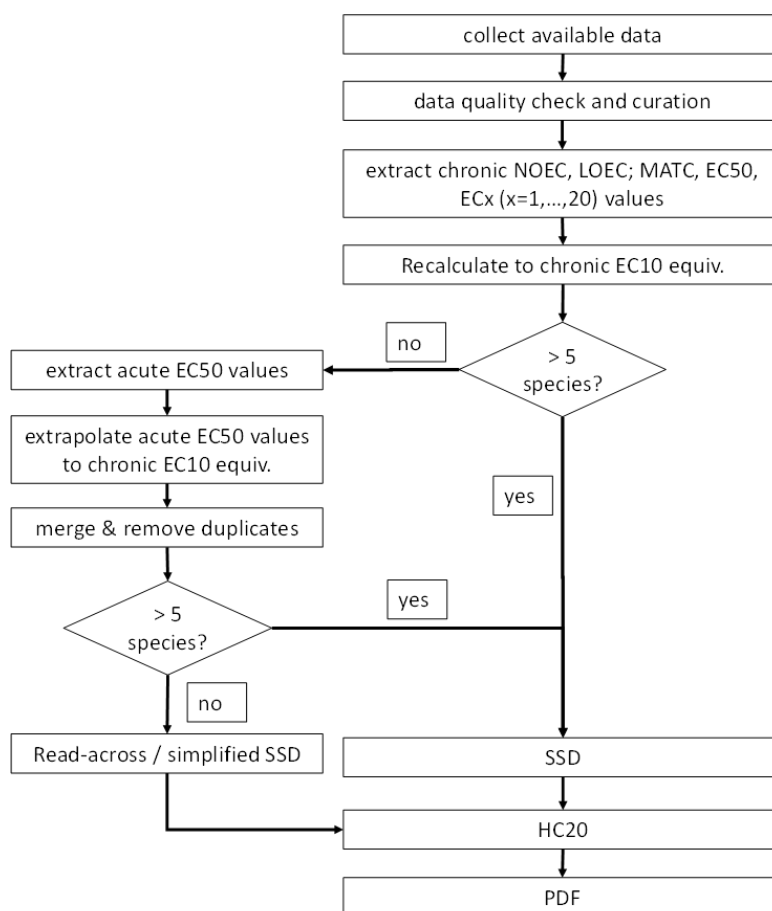
Source: UNEP- SETAC (2018) Pellston 2018 draft report

In case the minimum number of species deemed necessary to draw a reliable SDD curve is not attained (n = 5), even with the use of acute EC50 and Chronic EC50 extrapolated to EC10_equivalent, a read across / simplified SSD is proposed to ensure that as many as possible substances have an hazard value. However, the methodology to apply for this 'read across / simplified SSD' has not yet been agreed.

4.4.1 EC-JRC-2018 substances HC20

EC-JRC implemented as far as possible the Pellston recommendations using the ECHA, EFSA and PPDB database to meet the EU Commission deadline of November 2018.

In parallel, the Pellston task force is currently merging different ecotoxicity database to increase the likelihood of substances having at least five species (ideally from three main trophic levels: producer – Algae; consumer - Crustacean, and predator - Fish).

From the data extracted from the REACH-IUCLID, OpenFoodTox and PPDB database, the following operation have been perform to derive final substance HC20 values:

Step 1: All data toxicity points were extrapolated to chronic EC10 equivalent using the following extrapolation factors (extrapolation factors were proposed at Pellston workshop and may be refine in the future)

— Acute EC50 * 0.1 -> tagged as 'Extrapolated from Acute EC50'

— Chronic EC50 * 0.3 -> tagged as 'Extrapolated from Chronic EC50'

— Chronic (LOEC, ECr10, EbC10, EC10, LC10 , EC16, EC20 , NOEC, LOEC, MATC, GM-MATC, ChV) * 1 -> tagged as 'Chronic EC10eq'

Step 2: For all substances with >= 5 species, the HC20 was directly derived from the SSD curve.

The next steps are proposed to ensure that a HC20 value can be derived for all substances available in our database, even if the minimum requirement of 5 species was not reached. These next steps will likely be overwritten when the work undergoing under the UNEP-SETAC life cycle initiative umbrella is ready.

The number of species and number of trophic groups available to derive the HC20 is further used to define a quality score for each value (see next chapter)

Step 3: For substance having < than 5 and more than 1 species toxicity data, the procedure described in step 2 is applied.

Step 4: For substances having only 1 species toxicity value and considering that this value is equivalent to an HC50 (starting hypothesis), an extrapolation factor (ExF) is proposed to convert this HC50 into an estimated HC20.

The extrapolation factor (ExF) is derived from substances having at least 3 species. The following extrapolation factors are proposed (see also Figure 23 for the relation between HC50 and HC20 for those type of substances):

— Organic, ExF = 0.41 (sd = 0.21, VarCoeff = 197%, n = 2138).

— Inorganic ExF = 0.34 (sd = 0.18, VarCoeff = 189%, n = 435).

— element ExF = 0.30 (sd = 0.18, VarCoeff = 190%, n = 16).

— Organometallic ExF = 0.37 (sd = 0.19, VarCoeff = 193%, n = 59).

— Petroleum products ExF = 0.53 (sd = 0.25, VarCoeff = 210%, n = 154).

Due to the little number of results available for element and organometallic, these categories are grouped together with inorganic substances:

— Inorganic/element/organometallic ExF = 0.34 (sd = 0.18, VarCoeff = 190%, n = 510).

No significant differences were observed between mono-constituent, multi-constituents and UVCB substances (0.40, 0.43 and 0.41 respectively).

The estimated HC20 values for those substances were considered of low quality.

Figure 23: Relationship between HC50 chronic and HC20 chronic for various type of substances. Extrapolation performed only on substances having at least 3 or more species geometric means.
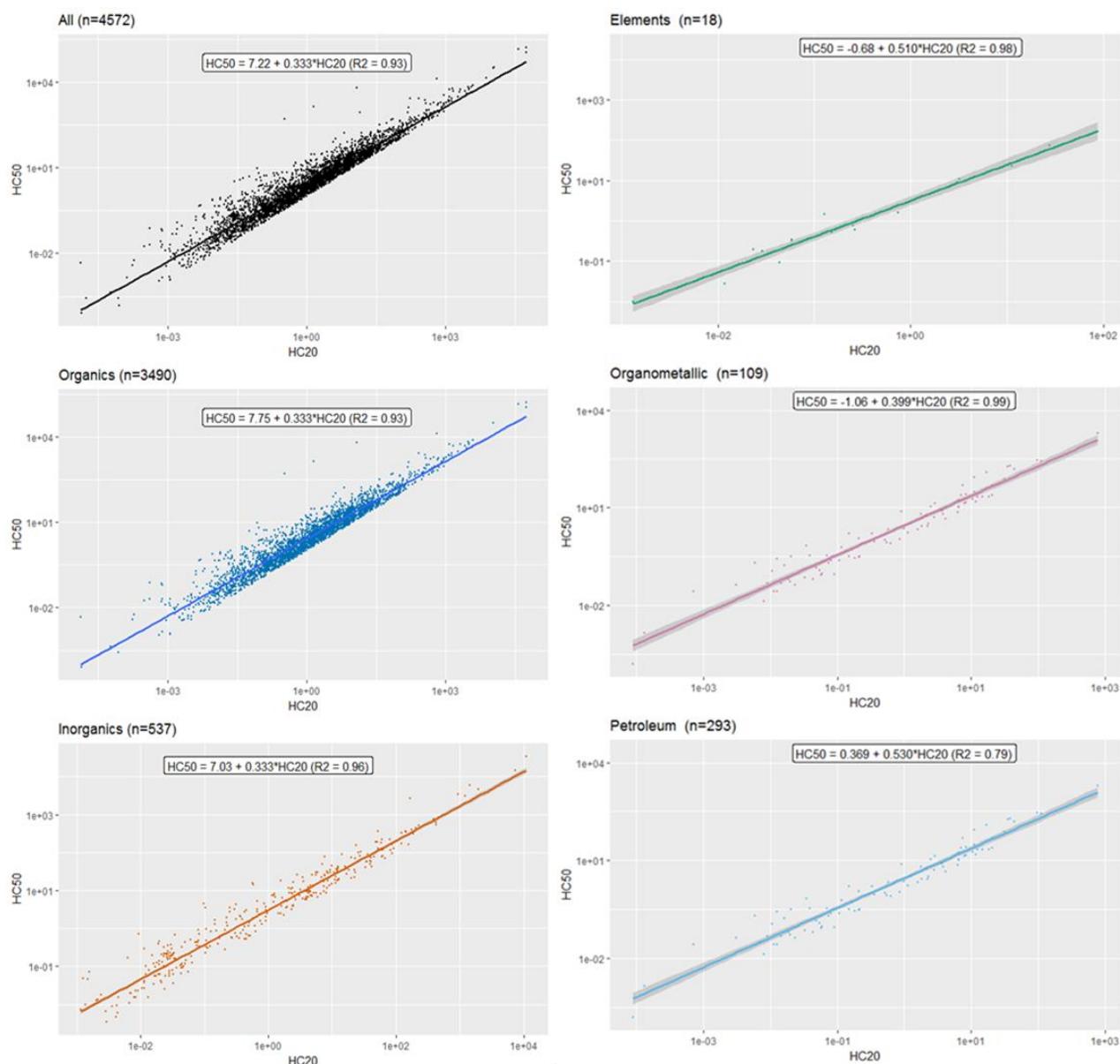
The total number of substance HC20 derived is 6764 from the joint REACH/EFSA database and 1316 from PPDB.

The availability per type of substance is presented in table 28. It should be stressed out that HC20 for metals are not recommended to be used with the USEtox® model as the procedure we applied did not follow the recommended USEtox® approach (toxicity values must be correct based on the free ion concentration).

Furthermore, HC20 values for UVCB, petroleum, inorganic, etc., were calculated but may not be used with current model before special attention was given to this type of substances.

Table 28: Number of substance HC20 based on SSD derived from EC10_eq per type of substances.

Composition	Type of substance	Number
Mono-constituent	Organic	3606
	Inorganic	566
	Element	21
	Organometaliic	92
	Petroluem	22
	No info/tie	94
	Total "mono"	**4401**
Multi-constituent	Organic	506
	Inorganic	36
	Organometaliic	8
	Petroluem	3
	No info/tie	16
	Total "multi"	**569**
UVCB	Organic	1202
	Inorganic	102
	Organometaliic	44
	Petroluem	306
	No info/tie	96
	Total "UVCB"	**1750**
No info/tie	Organic	34
	Inorganic	4
	Organometallic	1
	No info/tie	5
	Total "No info/tie"	**44**
TOTAL		**6764**

The number of substances for which at least five species are available to draw an SSD is rather limited (Table 29).

Table 29: Number of species available to draw the SSD curve and corresponding number of substances

Number of species to draw SSD	Number of HC20 from ECHA/EFSA	Number of HC20 from PPDB
1	1903	311
2	1544	678
3	1521	840
4	712	804
5	349	670
6	239	222
7	96	98
8	86	
9	26	
10	22	

Number of species to draw SSD	Number of HC20 from ECHA/EFSA	Number of HC20 from PPDB
11	28	
12	15	
13	8	
14	4	
15	9	
16	36	
17	21	
18	13	
19	5	
20	11	
21	2	
22	1	
23	3	
24	4	
25	14	
26	7	
27	10	
28	10	
29	4	
30	17	
31	6	
32	2	
33	2	
34	1	
35	8	
36	3	
37	2	
44	1	
45	1	
50	2	
51	1	
52	1	
54	1	
63	1	
70	1	
71	1	
72	1	
73	9	
Total	**6764**	**1316**

Less than 50% of the HC20 were derived with datasets having at least the three main trophic levels: Algae, Crustacean and Fish (Table 30).

Table 30: Number of substance for which the HC20 has been derived with a dataset including at least the 3 main trophic levels: Algae, Crustacean and Fish.

Having at least Fish, Crustacean and Algae	Number of HC20 from REACH/EFSA	Number of HC20 from PPDB
No	3840	812
Yes	2914	504
Total	**6764**	**1316**

HC20 were derived from a REACH/EFSA database composed for about 57% of EC10-eq values (NOEC, LOEC, EC10, etc.), 42% of acute EC50 and 1% chronic EC50 extrapolated (see table 31). Different situation regarding PPDB, being Chronic EC50 equivalent endpoints missing. The large majority (71%) was extrapolated from acute EC50, whereas only 29% is Chronic EC10 equivalent (NOEC) (Table 31).

Table 31: Number of toxicity value based on EC10_eq and based on Acute EC50 and Chronic EC50 extrapolated to EC10_eq.

Type of data	Data from ECHA/EFSA		Data from PPDB	
Chronic EC10_eq equivalent	13507	57%	1201	29%
Chronic EC50 equivalent	366	1%		
Acute EC50 equivalent	9911	42%	2893	71%
Total number of results	**23784**		**4094**	

4.4.2 Proposed quality score

To help EF practitioners to appreciate the level of 'reliability' of the data, a quality score (QS) is associated with each HC_{20} (equation 6).

Three criteria were used to distinguish from low to high quality HC_{20}.

1. Number of species available
2. Number of taxonomic groups (with best score if Algae, Crustacean and Fish are present)
3. Number of data extrapolated

Equation 6

$$QS = \ln(species) \times \ln(taxonomic\ groups) \times \frac{1}{(1 + extrapolations)^{0.1}}$$

The logarithmic function is selected because it allows reducing the relevance of large numbers (e.g. lots of species available or taxonomic groups) on the QS. Three taxonomic groups (algae, crustaceans and fish) and five different species are necessary to guarantee ecological realism. Once these criteria for these parameters are met, it is less important

"how much" such criteria are satisfied, hence a higher number of species or taxonomic groups would have a relative small influence on the QS.

The contribution of the extrapolation has a lower relevance on the QS compared to number of species and taxonomic groups. The 0.1 exponent was arbitrarily chosen to produce a decreasing curve not strongly penalising results with high number of extrapolations since extrapolated data might be anyway considered of good quality.

The numerical quality score can be converted in a qualitative assessment using the following thresholds (Table 32):

— $0 \leq QS < 1.48 \rightarrow$ LOW QUALITY (SSD with limited reliability)

— $1.48 \leq QS < 1.77 \rightarrow$ INTERMEDIATE QUALITY (reliable SSD)

— $QS \geq 1.77 \rightarrow$ HIGH QUALITY (highly reliable SSD).

Because: QS = 1.77 corresponds to the situation with 5 species, 3 SSD groups and no extrapolations and QS = 1.48 represents the situation with 5 species, 3 SSD groups but all values extrapolated.

The proposed quality scoring method avoids misclassifying cases in which outliers are present. For example, a LOW quality score would be assigned to a substance for which a high number of species are available belonging to the same taxonomic group. This hampers the ecological realism.

Table 32: Number of substances with a hazard value (HC20) and associated quality score.

Quality score	Number of substances	QS values
LOW	5591	$0 \leq QS < 1.48$
INTERMEDIATE	530	$1.48 \leq QS < 1.77$
HIGH	643	$QS \geq 1.7$
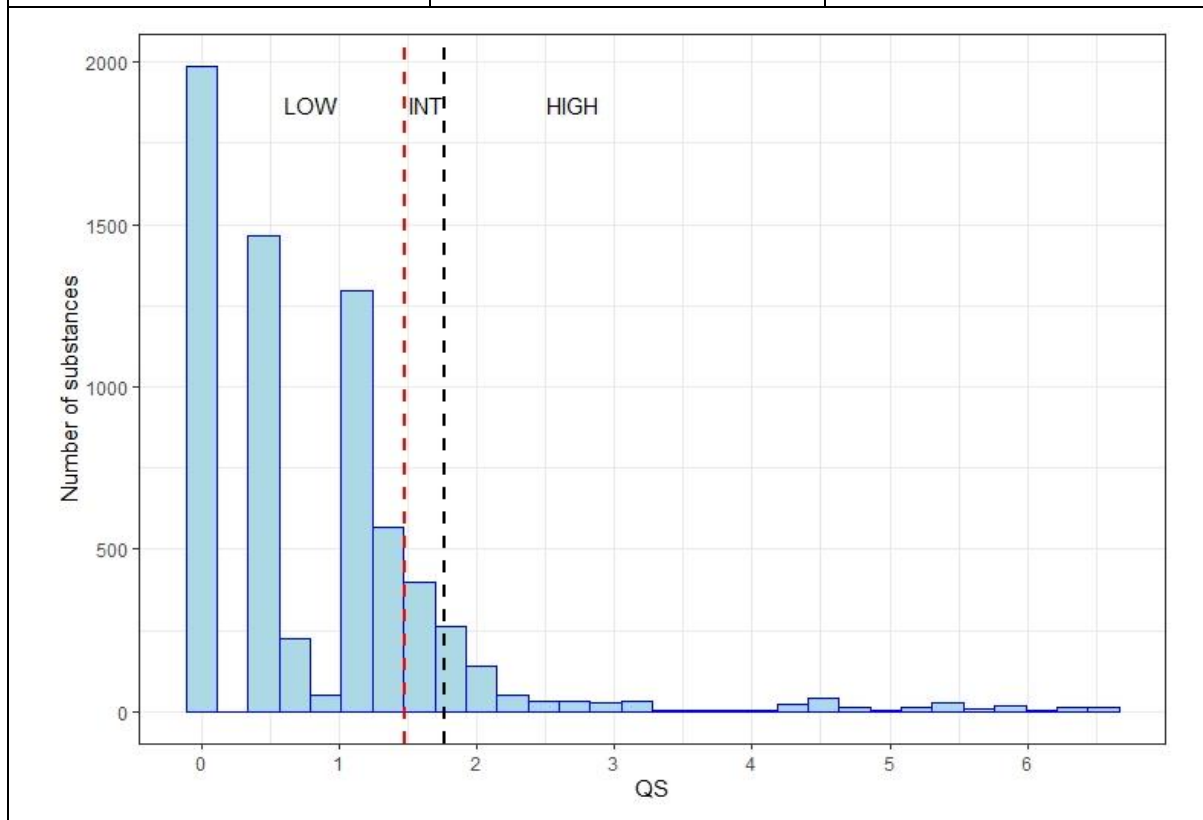
5 Deriving substance human toxicity data from the REACH and EFSA database

The R code used to retrieve human toxicity data from the REACH-IUCLID, OpenFoodTox and PPDB database and the list of variables available for each parameters are available in the online supplementary material (see annex 1 for list and http://eplca.jrc.ec.europa.eu/LCDN/developerEF.xhtml). However, the final individual species toxicity values are not available on the online materials for property and confidentiality reasons, but there are available on the ECHA dissemination website. Only the final substances hazard values are made available.

5.1 Human toxicity cancer

The REACH-IUCLID and OpenFoodTox database does not contains human toxicity value for cancer in a format compatible with the USEtox® method, therefore the human toxicity cancer effect factor has not been changed. For substances for which new physicochemical data were available, new fate and exposure factors were calculated.

5.2 Human toxicity non-cancer from REACH-IUCLID

All the repeated dose toxicity (RDT) data present in the REACH database as of May 2015 were exported by ECHA from the IUCLID database.

Table 33 displays the total number of substances for which study reports and test results are available, while table 34 presents the type of substances included in the database.

Table 33: Number of substances, ESR and human toxicity results extracted from IUCLID for ingestion and inhalation route.

Exposure route	Number of substances	Study reports	Test results
Ingestion	5700	18474	28440
Inhalation	1947	8595	12941

Table 34: Composition of the database according to the composition and type of substances.

Composition	Type of substance	Ingestion	Inhalation
Mono-constituent	Element	20	14
	Inorganic	522	312
	Organic	2779	813
	Organometallic	74	28
	Petroleum produt	14	12
	Not specified	143	28
Multi-constituent	Inorganic	32	21
	Organic	442	66
	Organometallic	6	1
	Petroleum produt	2	3
	Not specified	18	4
UVCB	Inorganic	107	91
	Organic	1095	196
	Organometallic	38	10
	Petroleum produt	286	309
	Not specified	88	28
Not specified	Inorganic	3	2
	Organic	28	7
	Organometallic	1	0
	Not specified	2	2
	TOTAL	**5700**	**1947**

Several specific fields included in the RDT endpoint study records, were used to define selection criteria for non-cancer human toxicity effect. In particular: reliability, purpose flag, study type, test guideline, GLP compliance, species, duration of exposure, route of administration, etc.

Generally, the majority of information were entered by registrants using IUCLID predefined fields. However, a drop-down list of predefined options is often paired with a free text option in order to allow registrants to add details. Such "free text" fields represent one of the principal challenges when dealing with REACH data. As a matter of fact, the same information can be retrieved in more columns and, consequently, some pre-processing operations were needed in order to adjust tables. Nonetheless, the final effect value per substance derived from REACH data based on the developed criteria, coincided with the critical endpoint value in each study report in the majority of cases, which is often reported in the conclusion column (the most heterogeneous "free text" field). Furthermore, the main causes for observed discrepancies between the automatically selected value and the critical endpoint value were investigated, to understand eventual weaknesses of the method.

An identical workflow was followed for both oral and inhalation toxicity; due to the nature of the endpoint, ad hoc rules were set for the few differences characterizing the endpoint.

To facilitate the use of data, some pre-processing operations were performed. These include:

— Adjustment of tables by merging information coming from "predefined options" and "free text" columns

— Selection of standard endpoints based on test guidelines (e.g. semi-chronic from OECD TG 409)

— Conversion of values to a uniform unit of measure

— Curing of variables regarding species (e.g. rat, mouse, dog, etc.)

— Reference point (e.g. NOAEL, LOAEL, etc.) and type of endpoint (e.g. chronic, semi-chronic and sub-acute) was fundamental; as for the derivation of ED50s for human beings conversion factors (CnF) based on these information are mandatory.

5.2.1 Dealing with species

For most repeated dose toxicity studies after oral and inhalation exposure, rodents (rat specifically), are listed as the preferred species. However, many other species were used to generate experimental results, such as monkey, dog, rabbit, guinea pig and hamster.

Regarding non-rodent species, dog is the preferable species for most of the guidelines, with the exception of the delayed neurotoxicity ones, where only laying hens are recommended. Therefore, the order preference for non-rodents might be variable. For instance, dog and swine (if dog is not available) are suggested by OECD 409 (OECD, 1998), whereas primates are not recommended. It might occur to have two or more species reported in the same result (e.g. rat and dog). In these cases, in order to automatize the procedure, it was considered only the species more "similar" to human beings; with the following order: monkey, dog, cat, rat, mouse, rabbit, cattle, sheep, hamster, guinea pig and gerbil.

The list of species used and the relative number of records they are associated to is reported in table 35.

USEtox® methodology was followed and the proposed extrapolation factors were used to derive human health effects for oral exposure from animals (Fantke et al., 2017). When dealing with inhalation exposure the extrapolation factor for interspecies conversion is 1. Thus, where inhalation data are concerned, air concentrations for animals and human are generally compared directly. This approach implies standardization of inhalation data with reference to respiratory rates.

Table 35: List of species used, and number of experimental results available

Species	Oral toxicity	Inhalation toxicity
cat	25	13
cattle	21	1
dog	1572	228
ferret	49	
goat	1	
guinea pig	32	253
hamster	21	107
chicken	11	
human	31	8
miniature swine	1	6
mink	4	
monkey	140	227
mouse	2337	1537
pig	510	203
rabbit	50	8
rat	163	170
sheep	23404	10180
Not specified	68	
Total	**28440**	**12941**

5.2.2 Dealing with reference point

Based on the USEtox® methodology for the derivation of non-cancer ED50 values, only reference points defined as NO(A)EL, NO(A)EC, NEL (No Effect Level), LO(A)EL, LO(A)EC, LD50, LC50 should be selected.

Chronic descriptors should always be prioritized. Reference points were gathered in four main groups associated with the following order of preference:

— NOAEL group, which includes NOAEL, NOAEC, NOTEL and NAEL

— NOEL group, consisting of NOEL, NOEC and NEL

— LOAEL group, which comprehends LOAEL and LOAEC

— LOEL group, including LOEL and LOEC.

— LD50 and LC50.

Table 36 illustrates the number of results for each harmonized reference point. Acute qualifiers (LD50 and LC50) will not be considered hereinafter, since the few observations in which they occur do not meet any of the quality requirements.

Table 36: Number of results for each harmonized reference point

Oral exposure		Inhalation exposure	
Reference point	**Number of results**	**Reference point**	**Number of results**
NOAEL	16707	NOAEL	7156
NOEL	4060	NOEL	945
LOAEL	4277	LOAEL	2315
LOEL	518	LOEL	295
No reference point	2878	No reference point	2230
Total number of results	28440	Total number of results	12941

Grouping similar reference led to a significant simplification of the workflow programming. It should be noticed that this rule has been refined during the methodology testing. Initially

only two classes were considered (NOEL and LOEL), where no distinction was posed between N(L)OAEL and N(L)EOL values. However, this categorization led to a frequency of errors of about 11.8% in the HIGH quality oral toxicity dataset. Thus, it was decided to improve it reducing the frequency of errors to 8.5%.

Due to the presence of a free text field, many other reference point can occur in the database: "BMD", "concentration level", "dose level", "Maximum Tolerated Dose (MTD)", "Tolerable Daily Intake (TDI)", "OEL", "LC100", "LD5", "LD50/40", etc.

Toxicological results associated with these reference points were not considered; in table 36, these are grouped under the "No reference point" class. Harmonized reference points were assigned to 25425 and 10367 observations for the oral and inhalation exposure databases, respectively. The lists of not harmonized reference points and their relative frequencies are reported in online materials.

5.2.3 Dealing with USEtox® endpoint categories requirements and relative information in REACH-IUCLID database

Three test duration endpoint are applicable in USEtox®: chronic, semi-chronic and sub-acute (Table 37).

Table 37: Test duration for each endpoint category according to USEtox® documentation.

USEtox® endpoint category	USEtox® time test duration
Sub-acute	14-28 days (2-4 weeks)
Semi-chronic	29-210 days
Chronic	> 219 days

In REACH-IUCLID database, indication of exposure duration could be found in two columns: duration of exposure/treatment and endpoint. The first is a free text column and therefore complicate to automatize, while the latter contains information similar to USEtox® endpoint category. Unfortunately, it was frequently noticed a disagreement between information reported such columns or with what stated in the conclusion. Consequently, it was decided to use a more robust source and to assign an endpoint category according to the test guideline followed.

In fact, the experimental protocol should define unambiguously the duration and thus the endpoint class. Furthermore, guideline is a widely available information in the REACH-IUCLID database.

The list of all available guidelines in REACH-IUCLID database and the related endpoint category, assigned according to its standard test duration are available in the online supplementary material (see annex 1 for list and http://eplca.jrc.ec.europa.eu/LCDN/developerEF.xhtml).

Generally, each guideline is associated with an unequivocal duration; however, few guidelines are less specific. For each results associated with such guidelines, a default endpoint category was initially assigned, and subsequently the column duration of treatment/exposure was manually checked to verify the correctness of the automatic attribution. For instance, OECD TG 409 (Repeated dose 90 days oral toxicity in non-rodents (OECD, 1998) is strictly associated with semi-chronic studies, whereas OECD TG 422 (Combined repeated dose toxicity study with the reproduction/developmental screening test (OECD, 2015) is initially associated with a semi-chronic endpoint to be confirmed with a secondary manual check of the duration.

For both routes of exposure, two groups of experimental protocols were recognized: the OECD/EU/US EPA set, containing the more reliable guidelines and therefore used to characterize the highest quality level, and a second group with additional or former guidance, considered for the second and third quality levels.

Harmonized endpoint categories were assigned to 17380 and to 7977 results for ingestion and inhalation, respectively; as illustrated in table 38.

Results were omitted when it was not possible to attribute a category. Despite its high relevance and realism, the number of chronic observations is significantly lower than the number of sub-acute and semi-chronic measurements. This reflects the complexity of long duration experiments as well as the REACH requirements for repeated dose toxicity studies.

Table 38: Number of results for each endpoint category, assigned according to the guideline.

Oral exposure		Inhalation exposure	
Endpoint category	**Number of results**	**Endpoint category**	**Number of results**
Chronic	1753	Chronic	1009
Semi-chronic	9916	Semi-chronic	4330
Sub-acute	5740	Sub-acute	3040
No endpoint category	11031	No endpoint category	4562
Total number of results	28440	Total number of results	12941

Lastly, few data on humans in the form of epidemiological studies, case reports or information from surveillance programs are included in REACH-IUCLID database. Since there is no standard guidance for human examinations, endpoint categories were assigned based on the duration. In case this information was not reported, the results were not considered.

5.2.4 Dealing with test values presented as ranges

In the original file values are presented in two columns displaying the low and the high value ranges. However, only one value need to be retained to be used in the USEtox® model. Each value is associated to a qualifier. The qualifier "empty" (default option) is considered equivalent to "=". The large majority of results have a numeric value only in the low value field (24849 and 10679 observations for oral and inhalation route, respectively). It is assumed that, when entering a result, the logic is to report it in the lower band. There are however situations where no value was entered in the lower band but in the higher (932 and 413 results for oral and inhalation exposure, respectively).

The lower value was always preferred; and as a result, when dealing with ranges the upper boundary was never considered. When no value was reported in the lower band, high values were used for the selection procedure.

Finally, values associated with unbounded qualifiers ("<" or ">") were never used. Nonetheless, if the corresponding upper qualifier is "≤", "ca." or "=", the upper numerical field was chosen. For example: considering the range 50<NOAEL≤300 mg/kg bw, the 300 mg/kg bw is selected.

5.2.5 Data selection and derivation of human health ED50 values

As a general rule, when more than 1 result is available substance, the lowest value is always selected (conservative approach). Then, the selection is based on priority orders using the endpoint category (chronic is preferred, then semi-chronic and chronic) and for the reference point (NOAEL is preferred, then NOEL, LOAEL and last LOEL) (Figure 24). Chronic tests and NOAEL results as the most realistic and relevant (Figure 24).

Finally, an additional rule was set for inorganic substances: metals should be preferred in a dissolved or element form, as recommend in USEtox® methodology (Fantke et al., 2017).

Information to identify inorganic substances were retrieve from a different REACH database, afterwards the column effect level based on was investigated to spot results obtained with dissolved or element inorganic compounds.

Figure 24: Data selection and conversion to ED50 workflow.
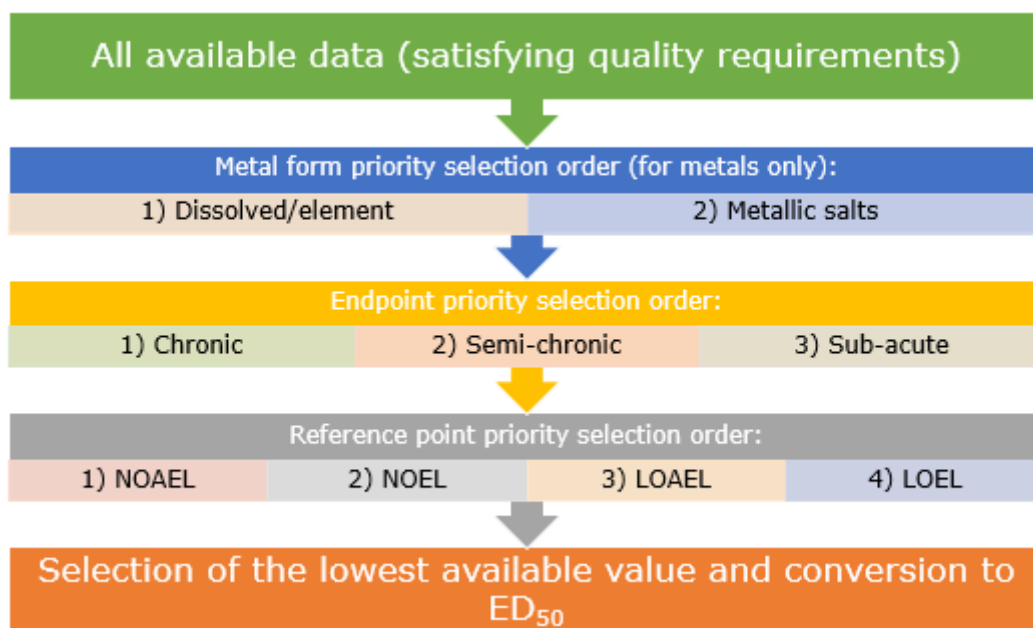
5.2.6 Quality levels

The criteria presented in Table 39 were used to assign three-quality levels to each test results retained for the final derivation of the substance non-cancer toxicity hazard value.

Each substance is associated only to a single quality level, representing the highest level assigned to at least one of its results. For instance, assuming that several results meeting HIGH and INTERMEDIATE quality levels criteria are reported for one substance, only the outcome derived from the HIGH level results was retained, as it represents the most reliable value. Table 40 reports the total number of test results available for each criteria.

Table 39: Criteria for quality level and number of results.

Criteria	HIGH	INTERMEDIATE	LOW
Reliability	K1 + K2	K1 + K2	K1 + K2
Adequacy of study	Key and supporting study + weight of evidence	Key and supporting study + weight of evidence	Key and supporting study + weight of evidence
GLP compliance	Yes	Yes, No, Not specified	Yes, No, Not specified
Guideline qualifier	According to, Equivalent to	According to, Equivalent to	According to, Equivalent to
Guideline	OECD/EU/US EPA	OECD/EU/US EPA Additional guidelines (Henkel method excluded)	OECD/EU/US EPA Additional guidelines (Henkel method included)
Route of administration (inhalation only)	Inhalation, gas, vapour, dust, aerosol, mist, fume	Inhalation, gas, vapour, dust, aerosol, mist, fume	Inhalation, gas, vapour, dust, aerosol, mist, fume

Criteria	HIGH	INTERMEDIATE	LOW
Inhalation exposure (inhalation only)	Nose, head, snout, face mask	Nose, head, snout, face mask, intratracheal, intralaryngeal, tracheotomy tube, inhalation chamber	Nose, head, snout, face mask, intratracheal, intralaryngeal, tracheotomy tube, inhalation chamber
Number of results (ingestion route)	4279	5352	4977
Number of results (inhalation route)	255	1695	1520

Table 40: Number of test results per criteria used to assign quality level (see table 39)

Reliability	Oral	Inhalat.	Adequacy of study	Oral	Inhalat.
K1	8738	5312	Key study	11421	5165
K2	14716	5578	Supporting study	9899	6222
K3	1837	1369	Weight of evidence	4048	783
K4	1836	579	Disregarded study	759	91
empty	1313	103	empty	2313	680
Study type			**GLP compliance**		
Experimental result	13942	7562	Yes, Yes (incl. certificate)	14064	5740
Read across	12841	4991	No	7551	2100
QSAR and calculation	114	99	No data, empty	6825	5101
Other types of information	185	64			
empty	1358	225			
Route of administration (inhalation only)			**Type of inhalation (inhalation only)**		
Inhalation		3854	Nose		1242
Gas		602	Head		215
Vapour		5055	Snout		2
Dust		585	Whole body		9351
Aerosol		2413	Face mask		1
Mist		4	Intratracheal		13
Fume		3	Intralaryngeal		3
Other routes or unspecified		312	Tracheotomy tube		1
Empty		113	Inhalation chamber		5
			Other types or unspecified		1360
			Empty		748

5.2.7 Conversion to non-cancer lifetime human health ED50 via ingestion and inhalation toxicity

Initially, ED$_{50chronic}$ values were derived using proper conversion factors (CnFs) to obtain chronic values from semi-chronic and sub-acute tests (Table 41), as well as to extrapolate ED50 values from NOAEL, NOEL, LOAEL and LOEL. The following equations were used, according to USEtox® methodology (Fantke et al., 2017):

Equation 7

$$ED_{50,chronic,oral}\left[\frac{mg}{kg\ day}\right] = \frac{REACH\ value\left[\frac{mg}{\frac{kg\ bw}{day}}\right] \times CnF_{reference\ point}}{CnF_{endpoint}}$$

$$ED_{50,chronic,inhalation}\left[\frac{mg}{m^3}\right] = \frac{REACH\ value\left[\frac{mg}{m^3}\right] \times CnF_{reference\ point}}{CnF_{endpoint}}$$

Table 41: Conversion factors for inhalation toxicity

Endpoint category	Conversion factor
Chronic	1
Semi-chronic	2
Sub-acute	5
NOAEL and NOEL	9
LOAEL and LOEL	2.5

After having derived non-cancer chronic ED50 for laboratory species; USEtox® model requires the following calculations to derive lifetime ED50 for humans via ingestion and inhalation exposure, respectively:

Equation 8

$$ED_{50,ingestion}\left[\frac{\frac{kg}{person}}{lifetime}\right] = \frac{\left(ED_{50,chronic,oral}\left[\frac{mg}{kg\ day}\right] \times bodyweight[kg] \times lifetime[year] \times 365\left[\frac{day}{year}\right]\right)}{CnF_{species} \times 10^6\left[\frac{mg}{kg}\right]}$$

$$ED_{50,inhalation} = \frac{\left(ED_{50,chronic,inhalation}\left[\frac{mg}{m^3}\right] \times inhalation\ rate\left[\frac{m^3}{day}\right] \times lifetime[year] \times 365\left[\frac{day}{year}\right]\right)}{10^6\left[\frac{mg}{kg}\right]}$$

Where the average bodyweight is 70 kg, the average lifetime is 70 years and the average inhalation rate is 13 m3/day. CnFspecies is the extrapolation factor for interspecies differences, its values are reported in table 42; it is always equal to 1 for the inhalation exposure.

Table 42: Conversion factors for interspecies differences.

Species	CnF interspecies
Human	1.0
Pig/swine	1.1
Dog	1.5
Monkey	1.9
Cat	1.9
Rabbit	2.4
Mink	2.9
Guinea pig	3.1
Rat	4.1
Hamster	4.9
Gerbil	5.5
Mouse	7.3

5.2.8 Route-to-route extrapolation

USEtox® suggests the use of route-to-route extrapolation in order to enlarge the number of available data (Fantke et al., 2017). Many studies have been performed to assess route-to-route feasibility (Dourson et al., 2001; Pepelko & Withey, 1985) and many extrapolation factors have been proposed to be used for different types of effect (Schröder et al., 2016). However, this procedure is recommended only in presence of verified systemic toxicity. Systemic toxicity is not often clearly indicated in the REACH-IUCLID database.

Despite this, in order to enlarge the human toxicity data, route-to-route extrapolation was applied. Each extrapolated number was divided by an arbitrary factor of 10 to cover the

uncertainty associated to the extrapolation. The number of available results for ingestion and inhalation exposure increased of 512 and 3410 units, respectively.

5.2.9 Results

Human non-cancer toxicity ED50 values were retrieved for **4523** substances (Figure 25 and table 43).

Figure 25: ED50 for ingestion and inhalation exposure for each quality level. Route-to-route extrapolation excluded
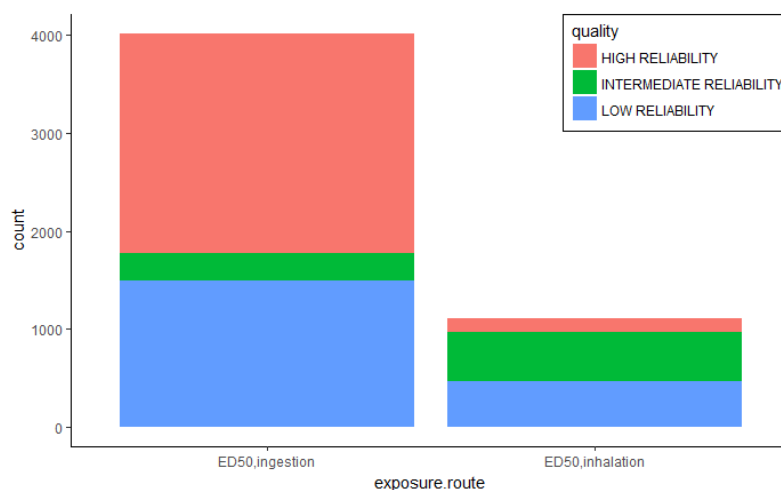
Table 43: Number of substances assigned to each quality level

Quality level	Oral Exposure	Inhalation Exposure
HIGH	2239	145
INTERMEDIATE	282	497
LOW	1490	471
Extrapolated from HIGH	81	1989
Extrapolated from INTERMEDIATE	264	238
Extrapolated from LOW	167	1183
Total from REACH data	4011	1113
Total extrapolated	512	3410
Total number of substances	4523	4523

The figure 26 shows the relation between the original USEtox® 2.1 values and the values generated with the REACH-IUCLID database.

A conformity check between the outcomes of the automated process and the conclusion reported in the ECHA was performed to verify that the approach apply on the REACH-IUCLID database led to the same outcome. For this purpose 250 and 94 results with high quality scores were considered for oral and inhalation endpoint, respectively.

For ingestion toxicity, only 16 discrepancies were found out of 187 results; whilst for inhalation toxicity, 3 discrepancies were spotted on a subset of 51. As a result, the analysis led to a relative frequency of incongruities of 8.5% and 5.8% for ingestion and inhalation toxicity, respectively.

The main reasons for eventual discrepancies are:

— Unit conversion based on the use of default factors for intake estimation. This type of mismatched was deemed acceptable.

— Presence of errors/incongruities in the study report, or at least in the fields considered in the automated process.

— Different study length than the time duration implemented in the system for a specific guideline. These differences might be caused either by different experimental choices, which cannot be controlled, or by different study length allowed by the less strict testing protocols (e.g. OECD TG 422).

Figure 26: Relation between REACH-derived and USEtox® ED_{50} values for ingestion exposure (top) and inhalation exposure (bottom) (n=596). The blue line represents the bisector.
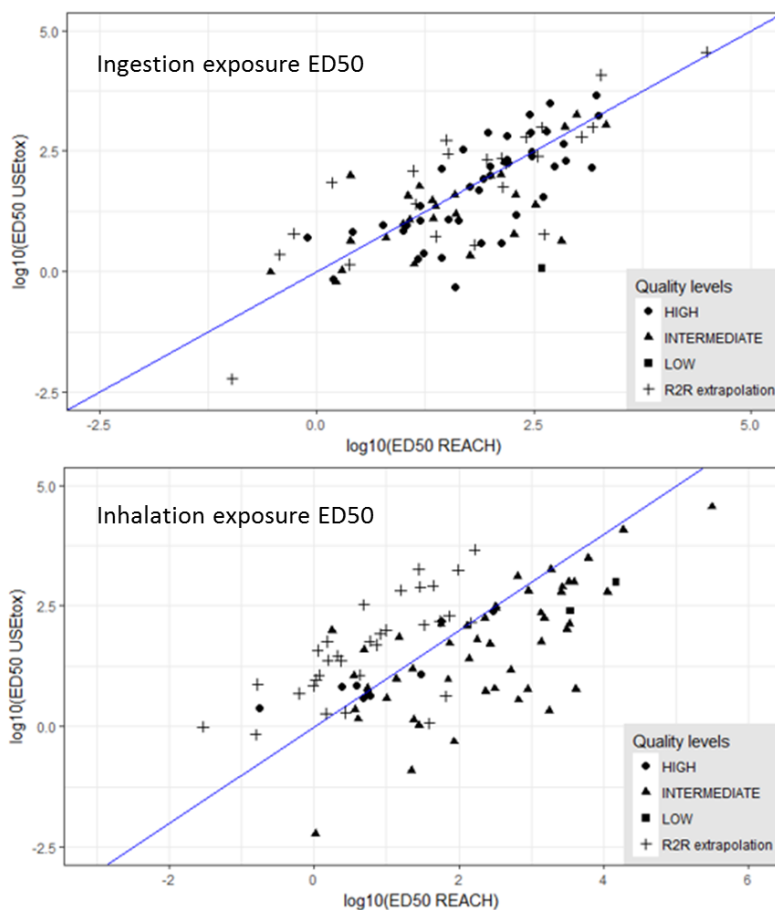
5.3 Human toxicity non-cancer from OpenFoodTox database

For human health, the EFSA toxicity data form the basis for the hazard and risk characterisation leading to either a health-based guidance value (e.g. ADI, AOEL, ARfD) or margin of exposure/safety values. In the case of animal health, relevant toxicity data for sensitive animals are also entered.

Health-based guidance values (HBGV), such as ADI (Acceptable Daily Intake), AOEL (Acceptable Operator Exposure Limit), ARfD (Acute Reference Dose) values, are also used for a decision about the approval of an active substance and in the context of the risk assessment and management process for the authorization of plant protection products.

These three hazard reference values can either be equivalent, or different for a substance, depending on the critical mammalian toxicity endpoints. Moreover, it may occur that one, or more, of these values cannot be derived for a substance.

When possible, during the selection procedure, data used to derive hazard reference values (ADI, AOEL, ARfD) were prioritized.

Because REACH-IUCLID and the OpenFoodTox databases serve different purposes, data contents and structure differ between the two databases. Therefore, also the data selection criteria vary between the databases:

— only the pesticide class is taken into account in the OpenFoodTox, with the exclusion of hazard data on food and feed additives, mycotoxins, food contact materials, flavourings, nutrient sources, nutrient/technological/zootechnical addivites;

— no tiered approach and no quality levels were outlined for the use of OpenFoodTox;

— information on the adopted testing guideline per endpoint record is scarce in OpenFoodTox;

— as with the REACH database, genotoxicity data could not be used.

The same workflow was followed for both oral and inhalation exposure data. Separation of such categories was performed according to the route of administration (when no route of administration was reported it was assumed to be oral).

5.3.1 Pre-processing of the data

Initially, due to the nature of the downloadable Excel file (information distributed on various tables and sorted for topics) a rearrangement of OpenFoodTox tables contents into a single user-friendly file was needed.

Then, as OpenFoodTox contains data regarding many toxicological and ecotoxicological endpoints, it was necessary to extract only data related to human/mammalian health (4574 observations). To this purpose, the column 'STUDY_CATEGORY' was filtered selecting only data labelled with Human health or Animal (non-target species) health. In the OpenFoodTox database, Human health refers to critical endpoint studies used for the derivation of HBGV values for a substance. Whereas, animal (non-target species) health refers to endpoint values for sensitive species, for which no HBGV values were derived from.

Afterwards, few data curing operations were performed:

— Conversion of measurement units. Guidance followed for the conversion of values to mg/kg bw/day (oral exposure) and to mg/m3 (inhalation exposure) are the same used for REACH database.

— Removal of data related to cancer effects. Similarly to the approach used for the REACH-IUCLID database, specific keywords such as tumour, tumorigenic, oncogenic, neoplastic, cancer, carcinogenic, carcinogenicity were searched in the column of the database describing the observed effect (EFFECT_DESC) to identify which results are associated with neoplastic histopathology. A proper code was developed to discriminate text expressing "hepatic carcinogenic effects were found" and "there was no incidence of carcinogenic activity".

  ● Moreover, the column named 'BASIS' was used to identify observations not relevant for this work by omitting rows where a histopathology neoplastic basis was stated.

  ● This procedure allowed to detect and discard 79 observations related to carcinogenic effects.

— Removal of values associated with unbounded qualifiers (">" or "<"). 1158 observations omitted.

To summarize, after few general data curing operation, the size of the OpenFoodTox database has reduced from 4574 to 3337 observations, composed mainly by oral exposure data (3256 observations) and, in a small part, by inhalation exposure data (32 observations).

The remaining pre-processing operations described hereinafter concern the fundamental parameters used to derive the Conversion Factors (CnFs). For this reason, these will be described in more details.

5.3.2 Dealing with USEtox® endpoint categories and test time duration

Information regarding test time duration are included in the column named 'TESTTYPE' in OpenFoodTox.

None of the available test types exactly refer to Repeated Dose Toxicity studies. In general, the wording repeated dose toxicity is also not specifically used in the EFSA's Conclusions on pesticides. Therefore, several test types have to be selected, in order to cover all possibilities.

Table 44 reports the harmonization of all the 'TEXTTYPE' possible entries in the USEtox® endpoint categories. The USEtox® test time duration was generally followed for the USEtox® endpoint assignation. However, when no test time exposure was available, the USEtox® endpoint was defined according to the EFSA's test type study reported in the OpenFoodTox database. The EFSA and USEtox®'s test time durations were almost equivalent.

Table 44: Harmonization of EFSA OpenFoodTox test types in USEtox® endpoint category.

EFSA OpenFoodTox TESTTYPE	EFSA OpenFoodTox test time duration	Harmonized USEtox® endpoint category	USEtox® test time duration
Short term toxicity	6-28 days (1 month)	Sub-acute	14-28 days (2-4 weeks)
Subchronic	27-90 days (1-3 months; 4-14 weeks)	Semi-chronic	29-210 days
Reproduction toxicity	Variable duration		
Chronic	> 90 days > 3 months; ≥ 1 year)	Chronic	> 210 days

Results obtained from experiments with a duration shorter than 14 days were excluded. When no information about experimental duration ('EXP_DURATION_DAYS') was reported it was assumed to be in agreement with the relative endpoint category ('TESTTYPE').

There is a large number of records without the test duration information available. Records with this field *empty* should be retained in the assessment, with the only exception for the Reproductive test type, due to a possible variability in exposure duration. *Reproduction toxicity* was considered to be equivalent to *semi-chronic*; however, when no experimental duration was reported it was omitted.

Table 45 reports the number of observations for each USEtox® endpoint category for both oral and inhalation exposure data. Considering the above described restrictions, 1709 and 16 observations were omitted for oral and inhalation exposures, respectively.

Table 45: Number of observation for each endpoint category, for both oral and inhalation exposures.

USEtox® endpoint category	Oral exposure	Inhalation exposure
Sub-acute	225	2
Semi-chronic	673	10
Chronic	685	4

5.3.3 Dealing with species

Only toxicological data based on mammals (rat, mouse, dog, pig and rabbit) were considered. Epidemiological/study with volunteers were omitted from this analysis. Bird species were also omitted as they are mainly associated with an ecotoxicological assessment.

Table 46 reports the number of observations for each species for both oral and inhalation exposure data. In total, 1016 observations of the oral exposure subset were omitted as not referred to mammals.

Table 46: Number of observations for each species, for both oral and inhalation exposures

Species	Oral exposure	Inhalation exposure
Rat	1169	28
Mouse	105	2
Dog	320	0
Pig	7	0
Rabbit	28	1
Human	3	1

5.3.4 Dealing with reference points

Similarly to the approach used with REACH-IUCLID database, reference points were gathered in five main groups associated with the following order of preference: the NOAEL group (i.e. NOAEL, NOAEC), the NOEL group (i.e. NOEL, NOEC), the LOAEL group (i.e. LOAEL, LOAEC), and the LOEL group (i.e. LOEL, LOEC). Then, if any LC50 and LD50 can be selected.

Other reference points, such as BMDL, BMDL05, concentration level, dose level, LC10, NOEDD, LDD50, were found in the database but were not retained in the assessment.

Table 47 reports the number of observations for each reference points, for both oral and inhalation exposure. 238 and 1 observations (for oral and inhalation exposure, respectively) were characterized by reference points not considered in this analysis.

Table 47: Number of observations for each reference points, for both oral and inhalation exposure.

Reference points	Oral exposure	Inhalation exposure
NOAEL	1671	23
NOEL	658	0
LOAEL	0	0
LOEL	1	0
LC_{50}	688	8

5.3.5 Data selection and derivation of human health ED50 values

The approach followed is very similar to the one applied for REACH data: when more values are available for each substance, the lowest value is selected in the final stage of the automated process, irrespective of the toxicological effects. Before applying this general rule, a priority order was established for the endpoint category (chronic, semi-chronic and last sub-acute) and for the dose descriptor (NOAEL, NOEL, LOAEL, LOEL and last LC50). This produced a decision tree that considers chronic tests and NOAEL results as the most relevant.

In addition, the only difference with REACH data, is the hazard reference value. An ad hoc rule was defined in order to prioritize values used in a hazard assessment context (Figure 27).

Measurement units for the selected effect level values were converted using equations and conversion factors as previously described for REACH data.

Lastly, data related to organic and inorganic substances were divided. For this scope, the column 'COM_TYPE' was investigated and inorganics were identified by the entries inorganic and metal.

Figure 27: Selection and derivation procedure of human health ED50 values.
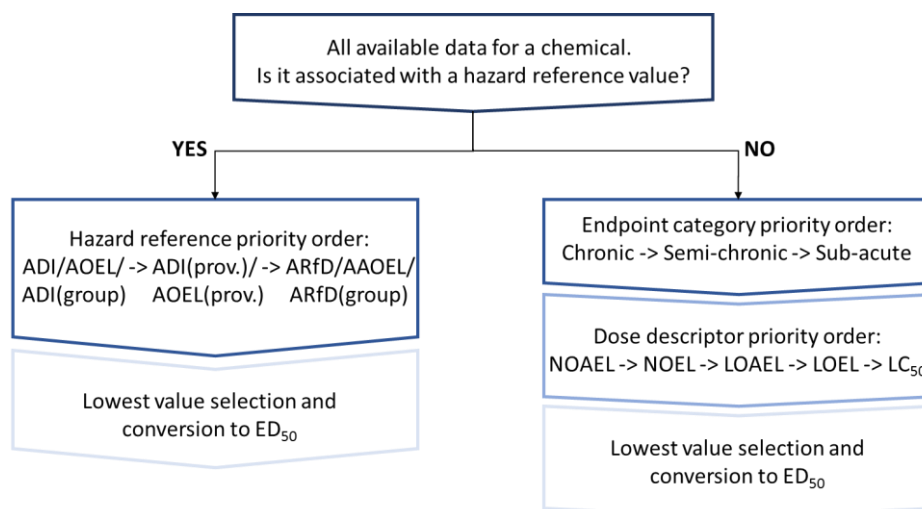
5.3.6 Pesticide with hazard reference values

In the OpenFoodTox various hazard assessment type ('ASSESSMENTTYPE') are reported, however only ADI, AOEL and ARfD were considered relevant for this analysis based on pesticides. Other assessment types, such as PNEC (Predicted No Effect Concentration), TDI (Tolerable Daily Intake) and MTDI (Maximum Tolerable Daily Intake) were not retained.

For each substance, data used to derive hazard reference values of interest (ADI, AOEL, ARfD) were preferred. Moreover, since more than one hazard reference values may be derived for the same compound, the following priority selection order was adopted, on the basis of the hazard reference value:

— Data used to derive ADI, ADI (group) or AOEL;

— Data used to derive ADI (provisional), AOEL (provisional) or AOEC (provisional);

— Data used to derive AAOEL, ARfD or ARfD (group).

The lowest risk value ('RISKVALUE_MILLI') was selected among those of the same selection level. For example, if for a substance both ADI and AOEL were derived the data associated to the lowest hazard reference value was selected. Generally, it was noted that ADI values were mainly associated with lower effect levels.

5.3.7 Results

For oral exposure, 437 and 18 ED50 were retrieved for organic and inorganic substances respectively.

For inhalation, 6 and 6 ED50 were retrieved for organic and inorganic substances respectively.

In general, it was noted an optimal agreement between the model output, and the matching critical endpoint study per substance in the OpenFoodTox database.

These values were then combined with the ED50 extracted from the REACH-IUCLID database.

6 New Characterisation factors for freshwater ecotoxicity and human toxicity

The final characterization factors calculated with the USEtox® 2.1 model, together with the list of new input parameters, are provided in the online supplementary material (see annex).

Important Note: Due to the ongoing work of the UNEP-SETAC life cycle initiative, building on the outcomes of the Pellston 2018 workshop, all the characterisation factors presented in this report may be eventually replaced by new ones. Both for the human and aquatic freshwater compartment, the Pellston workshop made important recommendations to improve the outcome of the model. Those recommendations require a significant intervention on the USEtox® model including how some input parameters are calculated.

For example, for the freshwater ecotoxicity impact category, the following recommendations were made:

— The substance hazard value to be based on 20percentile of a Species Sensitivity Distribution based on chronic EC10 equivalent reference points. This has been partially implemented in current EC-JRC-CFs (see chapter 4.4) but a new database on substance ecotoxicity is under construction which will lead to new substance hazard values and USEtox® effect factors.

— The bioaccumulation factor in the exposure factor equation to be removed (not implemented in current EC-JRC-CFs)

— A sediment compartment to be added to take into account the effect of substances that adsorbed on suspended particles and end-up in the sediment compartment (not implemented in current EC-JRC-CFs).

— Additionally, it was recommended that a terrestrial and a marine impact assessment category should be added to the model.

Characterisation factors for human non-cancer and freshwater aquatic toxicity PEF impact categories have been calculated using the USEtox® 2.1 model and using physicochemical and toxicity data from a variety of source.

The input table contains 3 tabs (Figure 28):

— tab 'Data input' contains all the input parameters for USEtox®. This tab is formatted to be used directly with the USEtox® model.

— tab 'Parameters quality' contains in addition the source and quality level of each value as well as the standard deviation when available.

— tab 'identifies' contains all possible names associated with each CAS / EC numbers with the smiles notation (Weininger, 1988) to help the identification of the substances.

Those input table is used in USEtox® to calculate the CF. The CFs table contains 2 tabs (Figure 29):

— The first 'tab' contains the CFs as calculated with USEtox® 2.1 model.

— The second tab specifies the source of the data.

CFs have been calculated for all substances available in the REACH/EFSA/PPDB database. Although the substances for which CFs have been calculated from EFSA and PPDB are either organic or inorganic substances, the REACH database contains a mix of substances of different composition and type: mono and multi-constituents, UVCB (Substance of Unknown or Variable composition, Complex reaction products, or Biological materials), and organic, inorganics, organo-metallics, elements, and petroleum products.

Figure 28: Screenshot of the excel file containing the input data to be used with the USEtox® 2.1 model.

Figure 29: Screenshot of the characterization factors excel file calculated with the USEtox® 2.1 model.

In order to report the characterisation factors calculated with USEtox® 2.1 in the Environmental Footprint reference package 3.0 compliant nomenclature, the emission compartments of USEtox® 2.1 had to be mapped to the emission compartment in the EF, as reported in Table 48.

Table 48: Equivalence between USEtox® and ILCD emission compartments.

	ILCD emission compartments	Equivalence with USEtox® emission compartments
Air	Emissions to air, unspecified	Average of urban/continental rural air
	Emissions to air, unspecified (long term)	0
	Emission to air, indoor	Average of Household/industrial indoor air
	Emissions to non-urban air or from high stacks	Continental rural air
	Emissions to urban air close to ground	Urban air
	Emissions to lower stratosphere and upper troposphere	Continental rural air
Water	Emissions to fresh water	Freshwater
	Emissions to sea water	Seawater
	Emissions to water, unspecified	Average of Freshwater/seawater
	Emissions to water, unspecified (long term)	0
Soil	Emissions to soil, unspecified	Average of Natural/Agric. soil
	Emissions to agricultural soil	Agric. soil
	Emissions to non-agricultural soil	Natural soil

The final list of characterisation factors is reported in the Annex, complemented with a column which inform the users regarding the source of data and the calculation principles underpinning the CFs (Figure 30). It consists of two tabs: in the first, CFs are related to ILCD emissions compartment and listed in rows. In the second, codes describing source of the data and the calculation principles are described.

Figure 30: Screenshot of the characterization factors excel file in the Environmental Footprint reference package 3.0 compliant nomenclature with the indication of the data source and the calculation principle.

76

6.1 CFs per type of substances

6.1.1 Organic substances

The USEtox® 2.1 model has been originally developed for mono-constituent organic substances. CFs calculated for all organic substances should be considered within the domain of applicability of the model.

CFs were calculated for all organic substances, disregarding their composition being, mono, multi constituents or UVCB. Organo-metallic substances have been modelled as organic substances.

6.1.2 Cationic metals (named 'inorganic database' as per USEtox®)

Although multimedia fate models were developed for organic substances (best suited), the USEtox® is also using this model to calculated CFs for cationic metals.

It should be noted that many aspects are currently not addressed in the fate modelling of metal via the USEtox® model:

— existing of a background concentration due to natural presence of metals in ecosystems

— essentiality of some metals for life, the fact that iron and zinc may actually be deficient in many ecosystems and for human, respectively

— the complex dynamic speciation of metals in the environment.

Until these specificities are not addressed, the potential toxicity impact assessment of metals in PEF context should be taken with high caution (see chapter on interpretation and weighting).

The following changes have been made by USEtox® to adapt the model to cationic metals:

— Fixing the degradation rate in water, sediment, soil and air to infinite time horizon (1E-22). Metals do not degrade.

— Octanol water partition used for organic substances are not suited for metals, therefore this value is set to '0' for metals.

— Instead, specific partition coefficient between water and 1) dissolved organic carbon, 2) suspended particles, 3) sediment particles and 4) soil articles are used.

USEtox® 2.1 provides CFs for 1 or 2 oxidative states for some metals. Those have been developed via a multi-years stakeholder collaborative effort and EC-JRC has decided not to change any of the input data agreed during this process.

Therefore, all CFs for the oxidative forms of the cationic metals (27 in total) provided with the USEtox® 2.1 model have been used as such. EC-JRC has made no modification of the input parameters.

However, since different form of metals can be listed in a product inventory output file, the following interventions have been made:

— When the elementary flow of a product inventory corresponds to the oxidative form listed in the USEtox® database, direct association was made.

— When the metal is reported as 'total metal' (Zinc, Copper, etc.) in the elementary flow of a product inventory (without any precision of the oxidative stage), the USEtox® proposed CFs was used.

— When two oxidative forms were available in the USEtox® database for the same metal, the average of the 2 values was used for the metal form, with exception for Chromium (see below).

— For Chromium, since the anthropogenic form of chromium emitted in the environment is only Ch(VI), the elementary flow reported in the inventory as 'Chromium' was associated with USEtox® CFs of Cr(VI).

— For human toxicity cancer, if USEtox® has reported that the oxidative form was carcinogen, the same value was used for the 'total' metal form.

6.1.3 Inorganic substances

Inorganic substance covers all substances that are not organic or organo-metallics.

Elements combined with an inorganic part (such as copper sulphate, zinc dichloride, aluminium hydroxide, Sodium sulphate, etc...) were treated as inorganics.

Like for cationic metals, the USEtox® model was used to calculated CFs with the same setting for biodegradability, Kow and sorption coefficients (see previous section).

In case degradation rate in water was available like for hydrogen peroxide that dissociates upon in water, the value was used as a surrogate of a 'biodegradation in water'.

Many other inorganics may undergo dissociation or degradation (photolysis in the upper part of the water column, hydrolysis, etc...) and requires special consideration for adapting the fate modelling to their specificity. Like for metals, the toxicity impact assessment should be interpreted with caution (see chapter on interpretation and weighting).

6.1.4 Group of substances and uncharacterized elementary flows

In the EF reference package (v.2.0) (EC-JRC, 2018b), there are about 97 elementary flows that appears in all 3000 datasets and for which no characterisation factors could be calculated using REACH, EFSA or PPDB database.

Several of these elementary flows are reported under a generic name such as PAH (Polycyclic Aromatic Hydrocarbons), PCB (Polychlorinated Biphenyls), VOC (Volatile Organic Carbons), NMVOC (Non Methane Volatile Organic Carbon), pesticides, fungicides, etc. or with names that do not allow identifying accurately the substance emitted. Others are reported with a CAS and/or EC number, but no data could be found.

When those flows are reported in product inventory with a mass (kg) emitted to water, air or soil compartment, their potential toxicity is not taken into account because no characterization factors is available.

To avoid that reporting emissions under a generic name are not contributing to overall toxicity score, a proxy is proposed to associate a CF to those emissions.

6.1.4.1 Elementary flows reported as 'Polycyclic aromatic hydrocarbons group', 'Volatile organic carbons, 'Non methane volatile organic carbons'

A weighted average (based on known global emission of individual PAH (Shen et al., 2013a)) of all the specific PAH substance CFs available in the EC-JRC-2018 database is used to allocate a CF to the group PAH (Table 49). The global emission of individual VOC and NMVOC were retrieved from the EDGAR database (Emission Database for Global Atmospheric Research) (EC-JRC, 2018a) (see online material).

For each emission compartment, a weighting factor was calculated for each flow for which a CFs in available in the EC-JRC-2018 database. The factor was then multiplied to the corresponding CFs and the sum of the weighted CFs was used as a proxy for the group PAH.

Table 49: Proposed calculation for allocating a CF to the group PAH reported in EF inventory.

CAS	Emission compartment	EC–JRC–2018 CFs ecotox	Global emission (kg)	Weighted factor	Weighted CFs
91-20-3	Emissions to fresh water	4.52E+03	2.30E+05	5.77E-01	2.61E+03
86-73-7	Emissions to fresh water	9.46E+03	1.80E+04	4.52E-02	4.27E+02
83-32-9	Emissions to fresh water	1.38E+04	3.20E+04	8.03E-02	1.11E+03
120-12-7	Emissions to fresh water	9.74E+05	1.00E+04	2.51E-02	2.44E+04
129-00-0	Emissions to fresh water	2.06E+06	1.90E+04	4.77E-02	9.82E+04
206-44-0	Emissions to fresh water	3.80E+05	2.40E+04	6.02E-02	2.29E+04
50-32-8	Emissions to fresh water	5.63E+04	3.50E+03	8.80E-03	4.94E+02
53-70-3	Emissions to fresh water	2.03E+04	1.80E+03	4.50E-03	9.20E+01
56-55-3	Emissions to fresh water	4.51E+06	7.20E+03	1.81E-02	8.15E+04
85-01-8	Emissions to fresh water	5.47E+04	5.30E+04	1.33E-01	7.28E+03
Total mass			3.98E+05	Weighted CF for PAH	2.39E+05

6.1.4.2 Elementary flows reported as 'adsorbable organic halogen compounds', 'Oils unspecified', 'Chloride', 'fungicides', 'herbicides', insecticides', and 'others'..

When the elementary flows refer to a group of substances for which global emission of individual substances are not available as for PAH, VOC and NMVOC (see above), an alternative approach was proposed with consist at using the 50%tile of all the individual CFs available in the EC-JRC-2018 database belonging to that group.

The table list the association made between reported elementary flows and the available CFs in the EC-JRC database.

6.1.4.3 Proposed characterization factors for group and other substances

The final CFs used for elementary flows reported in the reference package (v.2.0) under a generic name or for which not physicochemical and toxicity data were available in ECHA/EFSA and PPDB database are listed in table 50.

Table 50: CFs for aquatic toxicity and human non-cancer toxicity for elementary flows uncharacterized using the ECHA/EFSA/PPDB database with the proxy proposed.

Substance/Substance group	CAS number	Ecotox CF (emission to freshwater)	HH non cancer CF (emission to air, indoor)	Rule followed for its derivation
fungicides, unspecified		1.01E+05	1.24E-04	50th percentile of fungicides available
herbicides, unspecified		7.49E+04	4.64E-07	50th percentile of herbicides available
insecticides, unspecified		6.64E+05	1.78E-06	50th percentile of insecticides available
aldehydes, unspecified		6.97E+02	8.88E-08	50th percentile of aldehydes available
adsorbable organic halogen compounds		3.69E+03	1.23E-07	50th percentile of AOX available
chlorides, unspecified		3.01E+02	2.21E-08	50th percentile of inorganic chlorides available
chlorate	14866-68-3	3.01E+02	2.21E-08	50th percentile of inorganic chlorides available
chloride	16887-00-6	3.01E+02	2.21E-08	50th percentile of inorganic chlorides available
hydrogen chloride	7647-01-0	3.01E+02	2.21E-08	50th percentile of inorganic chlorides available
methylene chloride		3.01E+02	2.21E-08	50th percentile of inorganic chlorides available
hydrocarbons (unspecified)		8.01E+02	1.99E-09	50th percentile of petroleum products available
oils, unspecified		8.01E+02	1.99E-09	50th percentile of petroleum products available
chrysene	218-01-9	2.39E+05	7.85E-08	Average of PAHs available weighted on their emission amount
indeno(1,2,3-cd)pyrene	193-39-5	2.39E+05	7.85E-08	Average of PAHs available weighted on their emission amount
polycyclic aromatic hydrocarbons		2.39E+05	7.85E-08	Average of PAHs available weighted on their emission amount
methyl cyclopentane	96-37-7	3.62E+02	4.18E-08	Average of VOCs available weighted on their emission amount
volatile organic compound		3.62E+02	4.18E-08	Average of VOCs available weighted on their emission amount
non-methane volatile organic compounds		5.90E+02	2.18E-08	Average of NMVOCs available weighted on their emission amount
ammonium	14798-03-9	2.49E+03	1.56E-10	CF of ammonia
bromate		1.57E+02		CF of hydrogen bromide
bromide		1.57E+02		CF of hydrogen bromide
c12-14 fatty alcohol		2.31E+03	8.05E-10	CF of fatty alcohol C18
cis-2-pentene		4.76E+02		CF of pentene
trans-2-pentene		4.76E+02		CF of pentene
cyanide	57-12-5	2.65E+04	1.33E-07	50th percentile of hydrogen cyanide, sodium cyanide, potassium cyanide and calcium cyanide
fluoride	16984-48-8	2.04E+01	6.25E-07	CF of hydrogen fluoride
fluorine	7782-41-4	2.04E+01	6.25E-07	CF of hydrogen fluoride
hydrocarbons, aromatic		2.22E+04	3.53E-08	50th percentile of aromatic petroleum products available
hydrogen arsenide	7784-42-1	1.52E+03	4.21E-03	CF of arsenic
hydrogen iodide		3.75E+02		50th percentile of potassium iodide and sodium iodide

80

Substance/Substance group	CAS number	Ecotox CF (emission to freshwater)	HH non cancer CF (emission to air, indoor)	Rule followed for its derivation
iodide		3.75E+02		50th percentile of potassium iodide and sodium iodide
lead dioxide	1309-60-0	6.89E+01	5.37E-03	CF of lead
methyl bromide		2.61E+03	1.38E-07	CF of bromine
tin oxide		2.98E+02		CF of tin

6.1.5 Uncharacterized elementary flows

About 130 elementary flows that are reported in the reference package (v.2.0) remain uncharacterized because no physicochemical and toxicity data could be found in the three consulted database (REACH, EFSA and PPDB). The list of uncharacterized flows is available in the supplementary materials.

Some of those substances have been pre-registered in REACH but not registration dossiers have been provided suggesting that the interest to use those substances in EU has dropped (no dossier = no market). However, those substances could be used in other regions of the world; however, without a REACH dossier they cannot be imported on the EU market either.

Some of those substances have been listed in the REACH Annex III which cover substances for which a full annex VII dossier is required (full physicochemical properties plus aquatic toxicity an human toxicity data) and this despite being used a low tonnage < 10 tons / year). The basis for requesting full annex VII information is due to the hazard profile of those compounds being either very toxic for aquatic life or for human. Therefore, if there are really used in product LCA (it is uncertain why these flows have been reported) a characterisation factor should be provided.

This requires however significant additional investigation to search, assess and generate new CFs.

6.2 Human toxicity cancer impact category CFs

In total, **621** human cancer toxicity CFS are available (same number as in USEtox® 2.1).

Since no new cancer effect factor could be retrieved from the REACH and EFSA database, the CFs reported in the USEtox® input data were used. However, when new physicochemical properties data were available from REACH / EFSA or PPDB database, those parameters were used to recalculate fate and exposure factors using the USEtox® 2.1 model.

Two types of CFs are available:

— **Type 1**:   Effect, Fate and Exposure factors are from USEtox® 2.1 original input data. This concerns 14 cationic metals and 403 organic substances.

— **Type 2**: Fate and Exposure were calculated with REACH physicochemical data, while the effect values was taken from USEtox® 2.1 original input data. This concerns 204 organic substances.

Most of the CFs for human cancer toxicity are therefore identical with the USEtox® CFs (Figure 31). For the one calculated with new fate and exposure factor, the relation shows that all the new calculated characterization factors are within one order of magnitude with the USEtox® 2.1 CFs. They can be considered similar.

Figure 31: Scatter plot showing the relation between new JRC-2018 CFs and the current USEtox® 2.1 CFs for human cancer toxicity.

Differences USEtox CF vs JRC CF, cancer, emissions to indoor air
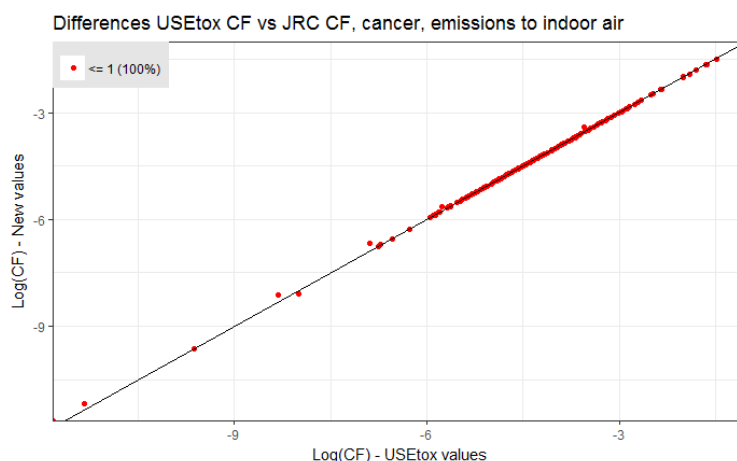
6.3 Human toxicity non-cancer impact category CFs

In total **3450** human toxicity non-cancer CFs are now available, compared to **426** originally available with the USEtox® 2.1 model. To ensure coverage of as many substances as possible, different database have been used and sometime combined to calculate final CFs. There are five 'type' of CFs:
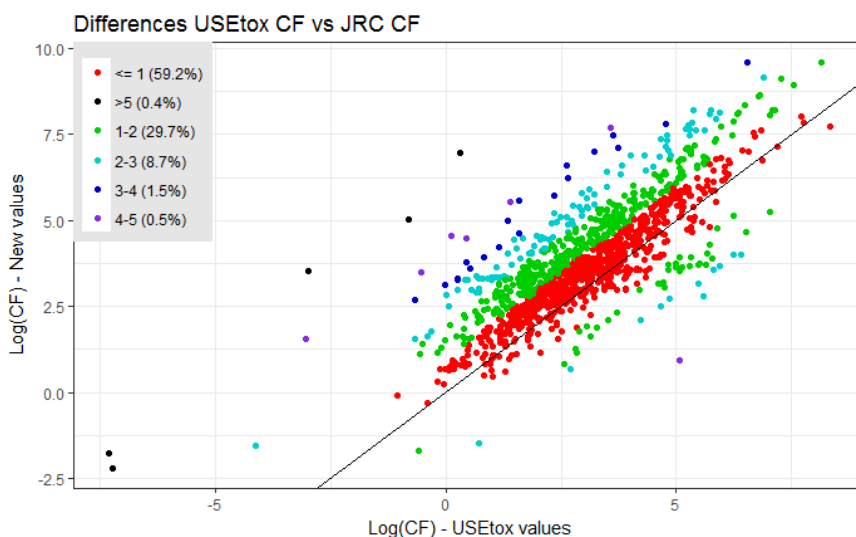
— **Type 1**: fate, exposure and effect factors are USEtox® 2.1 original input data. This concerns 27 cationic metals and 172 organic substances

— **Type 2**: Fate and Exposure were calculated with REACH physicochemical data, while the effect values was taken from USEtox® 2.1 original input data. This concerns 147 organic substances.

— **Type 3**: fate, exposure and effect factors were all calculated with REACH data. This concerns 2710 new substances and 106 substances that were listed in the USEtox® database

— **Type 4:** fate, exposure and effect factors were all calculated with EFSA data. This concerns 235 new substances

— **Type 5:** fate, exposure and effect factors were all calculated with PPDB data. This concerns 53 new substances.

For the majority of substances available in both the USEtox® database and in the EC-JRC-2018 database, CFs are within one order of magnitude (Table 51 and Figure 32). However, for few substances, CFs can be extremely different. No investigation was performed to understand the source of this variability.

Table 51: Ratios of the EC-JRC-2018 CFs and the current USEtox® 2.1 for human toxicity non-cancer presented by order of magnitude difference with total number substances in each bin and percentage, for emission to indoor air compartment.

| Order of magnitude difference	Number of CFs	Percentage
-3	7	2.5%
-2	12	4.3%
-1	100	35.8%
0	123	44.1%
1	11	3.9%
2	14	5%
3	9	3.2%
4	1	0.3%
5	1	0.3%
6	1	0.3%

Figure 32: Scatter plot showing the relation between new EC-JRC-2018 CFs and the current USEtox® 2.1 CFs for human non-cancer toxicity. Colour codes are used to distinguish the relationship per order of magnitude, for emission to indoor air compartment.
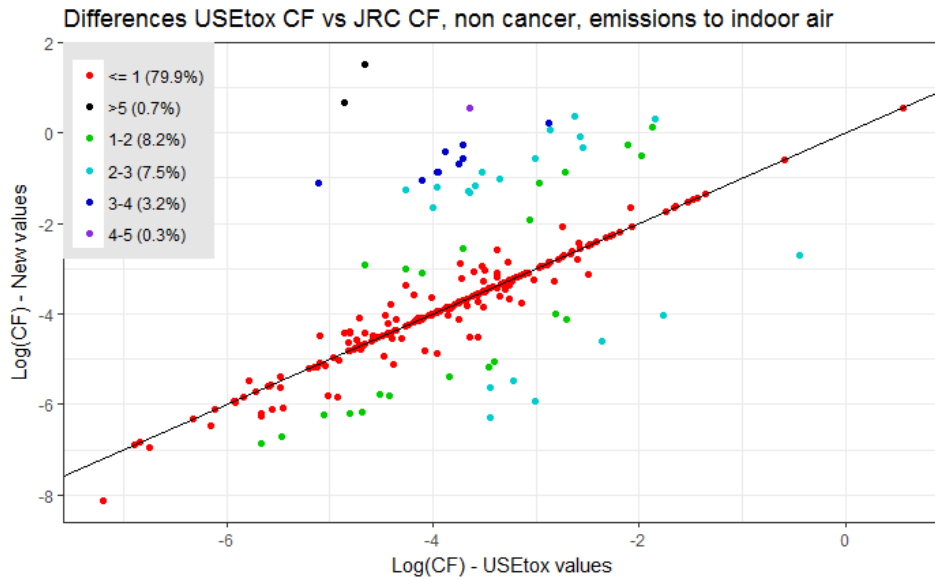
Characterisation factors are the results of a multiplication between fate, intake and effect factors. The figure 33 shows that the variability observed for the CFs can likely explained by the variability of the effect factor (substance hazard values) (bottom graph), since all the intake factors are less than 1 order of magnitude difference (top left graphs, red dots).

Figure 33: Scatter plot showing the relation between new EC-JRC-2018 CFs and the current USEtox® 2.1 CFs for human non-cancer toxicity CFs, Intake and Effect factors (both for inhalation and ingestion). Colour codes are used to distinguish the relationship per order of magnitude, for emission to indoor air compartment.
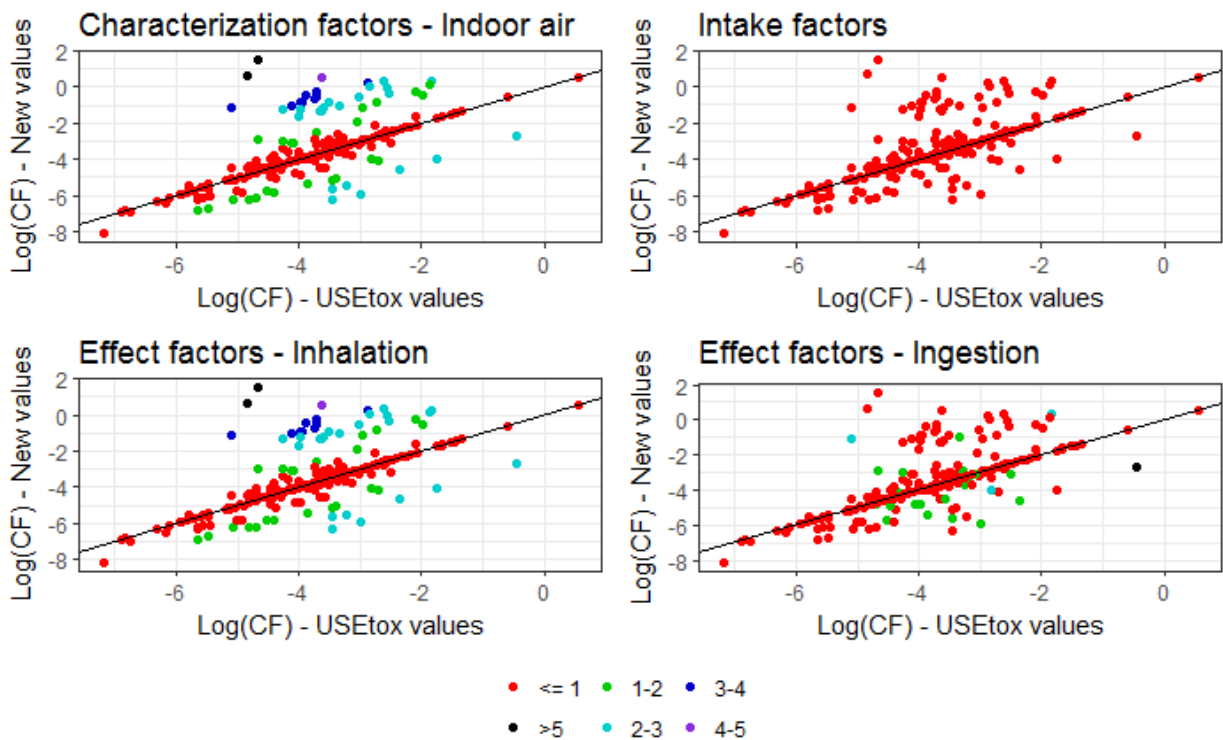
6.4 Freshwater aquatic toxicity impact category CFs

In total **6011** freshwater CFs are now available, compared to **2499** originally available with the USEtox® 2.1 model.

To ensure coverage of as many substances as possible, different database were used and sometime combined to calculate final CFs.

There are five 'type' of CFs:

— **Type 1**: fate, exposure and effect factors are from USEtox® 2.1 original input data. This concerns 27 cationic metals and 1230 organic substances

— **Type 2**: Fate and Exposure were calculated with REACH physicochemical data, while the effect values are from USEtox® 2.1 original input data. This concerns 191 organic substances.

— **Type 3**: fate, exposure and effect factors were all calculated with REACH data. This concerns 3006 new substances and 1078 substances that were listed in the USEtox® 2.1 original database.

— **Type 4:** fate, exposure and effect factors were all calculated with EFSA data. This concerns 289 new substances

— **Type 5**: fate, exposure and effect factors were all calculated with PPDB data. This concerns 190 new substances.

85% of the newly calculated CFs for emission to freshwater compartment are within 1 order of magnitude with the USEtox® 2.1 original CFs. However, few substances show very high differences (up to 5-6 order of magnitude) (Table 52 and Figure 34).

Table 52: Ratios of the EC-JRC-2018 CFs and the current USEtox® 2.1 presented by order of magnitude difference with total number substances in each bin and percentage, for emission to freshwater aquatic compartment.

Order of magnitude difference	Number of CFs	Percentage
-5	1	0.07%
-3	11	0.8%
-2	34	2.3%
-1	121	9.5%
0	631	48.7%
1	343	27.0%
2	99	7.8%
3	19	1.5%
4	6	0.5%
5	3	0.2%
6	2	0.1%

Figure 34: Scatter plot showing the relation between new EC-JRC-2018 CFs and the current USEtox® 2.1 CFs. Colour codes are used to distinguish the relationship per order of magnitude, for emission to freshwater aquatic compartment.

Characterisation factors are the results of a multiplication between Fate, Exposure and Effect factors. The figure 35 shows that the variability observed for the CFs can likely explained by the variability of the Effect factor (substance hazard values) (graph bottom left), since most of the fate and exposure factors are for a large part very similar (less than 1 order of magnitude difference – left 2 graphs).

Figure 35: Scatter plot showing the relation between new EC-JRC-2018 CFs and the current USEtox® 2.1 CFs for aquatic freshwater toxicity CFs, fate, exposure, and effect factors. Colour codes are used to distinguish the relationship per order of magnitude, for emission to freshwater compartment.

85
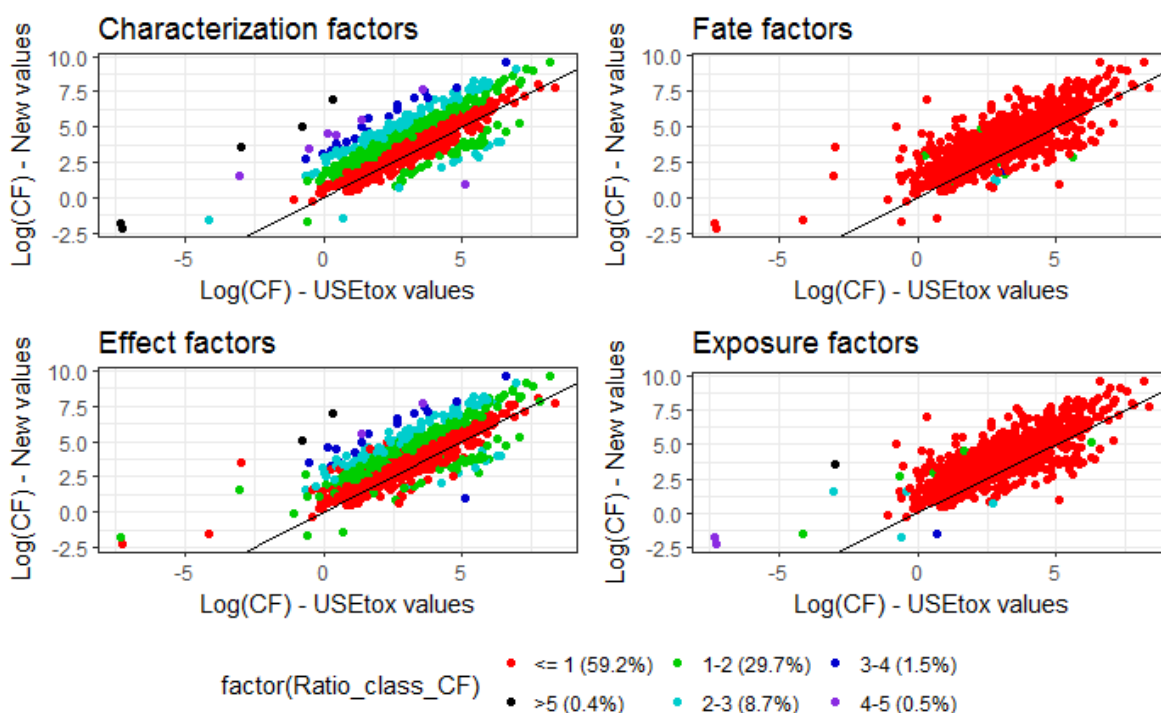
6.5 Final characterisation factors for organic, metals, essential metals and inorganic substances after robustness assessment.

As observed during the EF pilot phase on all the representative products, metals are usually dominating the impact score for human toxicity cancer and non-cancer, and to lesser extend also to the freshwater ecotoxicity score. This observation was also made on the LCA of more than 100 different types of products from food sector, construction, appliances and mobility (Castellani et al., 2018; Sala et al., 2018). Whatever product is analysed, metals always contribute the most to human toxicity. Recognizing that the USEtox® model is not yet covering all the unique properties of some categories of substances (multimedia have originally been developed for organic substances), a robustness assessment is proposed to take into account those specificities. Therefore, it is proposed to apply to each substance CFs by a default robustness factors reflecting both the appropriateness of the model but also the specificities of some groups of substances. The robustness factors applied on CFS and the justifications are described in table 53.

Table 53: Robustness factors applied to CFs

Group of substances	Robustness factors	Reasoning
Organics, organometallics, petroleum, UVCB	1	Multimedia fate modelling have been built for these substances (however, UVCB and organometallic included in this category may require still some special consideration)
Metals, non-essentials	0.1	Multimedia fate modelling are not best suited for metals. Furthermore, background concentration, essentiality, complex speciation, etc. are not yet taken into account in the fate and toxicity calculations.
Metals, essentials (Co, Cu, Fe, Mn, Mg, Mo, Se, Zn)	0.01	Multimedia fate modelling are not best suited for essential metals. For example, many ecosystems are deficient in iron (while the mass reported as emission in LCA is one the highest), or many humans are deficient in Zinc (while current LCA outcomes suggest zinc being the driver for toxicity.
Inorganics	0.1	Multimedia fate modelling are not best suited for inorganic. Furthermore, dissociation/degradation of inorganic is not yet taken into account in the model.

These robustness factors were directly implemented in the final CFs.

In order to allow the differentiation of impacts by "substance family", 3 sub-methods per main category are generated, namely:

— Human toxicity Cancer - metals

— Human toxicity Non-cancer - metals

— Freshwater ecotoxicity - metals

— Human toxicity Cancer - inorganics

— Human toxicity Non-cancer - inorganics

— Freshwater ecotoxicity - inorganics

— Human toxicity Cancer - organics

— Human toxicity Non-cancer - organics

— Freshwater ecotoxicity - organics

This approach is similar to the one adopted for the Climate Change impact category.

7 Contribution analysis

A contribution analysis analysing the EF representative products data sets (available at http://eplca.jrc.ec.europa.eu/EF-node/processList.xhtml?stock=EF_representative_data) was performed to evaluate the new sets of characterization factors, compared to the ones used during the EF pilot phase and included in the EF reference package 2.0.

The analysis was done with Look@LCI, a software developed by JRC, that was specifically designed to analyse EF-compliant data sets: indeed, it was used during the EF pilot phase as a reference tool to calculate the LCIA of the EF representative products (RP). By avoiding the step of mapping the EF nomenclature to a different one, all mistakes that occur during this step are avoided. This means that EF-compliant or ILCD-compliant data-sets can be analysed without need of implementing the files in existing LCA software.

The results delivered by Look@LCI are therefore to be considered the reference ones to be taken into account when analysing EF-compliant datasets: results that are not aligned to the ones of Look@LCI, calculated through other LCA software, are not EF-compliant.

Look@LCI allows the calculation of a large number of data sets, thus it is designed to analyse and check also full databases in a short time frame (> 1000 datasets/hour); all results are delivered as Excel files, each of them containing, among the others, the following information:

— LCIA results: characterized, normalized, weighted
— Contribution analysis of most relevant elementary flows within each impact category (relevance threshold to be decided by the user (default 80%))
— Contribution analysis of most relevant locations, for regionalized elementary flows.
— Summary of uncharacterized elementary flows
— Summary of processes with negative impact categories results
— Summary of processes with dominating impact categories (> 50%)
— — Within each impact category, it calculates the frequency of relevance of the elementary flows identified as "most relevant" in the data sets analysed. In practice, it identifies how often an elementary flow contributes to the relevance threshold (80%) (e.g. out of 100 datasets analysed, CO_2 is a most relevant elementary flow in 75 datasets. Maximum relevance, average relevance are also quantified).

7.1 Results of the contribution analysis performed on EF representative products

The analysis was performed using:

• the CFs used during the EF pilot phase (EF reference package 2.0[1]) and
• the new EC-JRC-2018 CFs, non-weighted and weighted (see previous chapter).

The detail results of the contribution analysis are provided in the online supplementary material (see annex)

7.1.1 Human toxicity, cancer

The elementary flows contributing up to 80% to the total impact score for human toxicity cancer are provided in table 54, for each representative product. The number of representative products for which a particular flow is relevant is reported. For example 'chromium emissions to fresh water' is relevant in all representative products, while 'mercury emissions to air, unspecified' is only relevant for 1 representative product out of 30. In addition, the maximum contribution of a given elementary flow is also reported.

---

[1] The EF reference package 2.0 is available at http://eplca.jrc.ec.europa.eu/LCDN/developerEF.xhtml

The following observations can be made:

— With EF 2.0, the most relevant elementary flow (expressed as unique substance) in the largest number of representative products is chromium. Cadmium, mercury and formaldehyde are also contributing but to a much lower extend (maximum contribution of 11% in 1 -3 representative products)

— EC-JRC-2018 without robustness factor on CFs reinforces the contribution of chromium. This is now the only substance that appears in the 80% contribution.

— EC-JRC-2018 with robustness factor on CFs provides a more diverse list of substances contributing to human toxicity-cancer, with metals but also organic and inorganic contributing to the score.

Table 54: Elementary flows contributing up to 80% to human toxicity cancer in all the 30 available EF reference products.

CFs from:	Elementary flow	max contribution	relevant in datasets	total datasets analysed
EF 2.0	chromium emissions to fresh water	79%	30	30
	chromium vi emissions to fresh water	58%	23	30
	chromium emissions to water, unspecified	50%	13	30
	chromium vi emissions to air, unspecified	27%	3	30
	chromium emissions to agricultural soil	16%	5	30
	chromium emissions to air, unspecified	20%	2	30
	cadmium emissions to agricultural soil	11%	3	30
	formaldehyde emissions to urban air close to ground	12%	1	30
	mercury emissions to air, unspecified	9%	1	30
	chromium emissions to non-urban air or from high stacks	8%	1	30
EC-JRC-2018 Without robustness factor	chromium emissions to fresh water	86%	29	30
	chromium emissions to non-urban air or from high stacks	36%	17	30
	chromium emissions to water, unspecified	36%	12	30
	chromium vi emissions to fresh water	36%	12	30
	chromium emissions to agricultural soil	22%	6	30
	chromium vi emissions to air, unspecified	33%	3	30
	chromium emissions to air, unspecified	37%	2	30
EC-JRC-2018 with robustness factor	chromium emissions to fresh water	32%	29	30
	benzo[a]pyrene emissions to non-urban air or from high stacks	17%	22	30
	formaldehyde emissions to air, unspecified	53%	17	30
	chromium emissions to non-urban air or from high stacks	15%	14	30
	benzo[a]pyrene emissions to fresh water	41%	14	30
	chromium vi emissions to fresh water	26%	14	30
	cadmium emissions to agricultural soil	49%	12	30
	mercury emissions to air, unspecified	32%	11	30
	chromium emissions to water, unspecified	20%	11	30
	chromium emissions to agricultural soil	11%	8	30
	nickel emissions to agricultural soil	6%	5	30
	nickel emissions to water, unspecified	20%	5	30
	2,3,7,8-tetrachlorodibenzo-p-dioxin emissions to fresh water	4%	4	30
	formaldehyde emissions to urban air close to ground	82%	4	30
	chromium vi emissions to air, unspecified	26%	3	30
	nickel emissions to non-urban air or from high stacks	5%	2	30
	polychlorinated biphenyls emissions to agricultural soil	23%	2	30
	nickel emissions to air, unspecified	9%	2	30
	benzo[a]pyrene emissions to air, unspecified	7%	2	30
	prochloraz emissions to agricultural soil	13%	2	30
	nickel emissions to fresh water	2%	1	30
	propylene oxide emissions to fresh water	4%	1	30
	chromium emissions to air, unspecified	16%	1	30
	arsenic emissions to non-urban air or from high stacks	5%	1	30

CFs from:	Elementary flow	max contribution	relevant in datasets	total datasets analysed
	lead emissions to agricultural soil	3%	1	30
	mercury emissions to agricultural soil	4%	1	30
	furan emissions to non-urban air or from high stacks	3%	1	30

7.1.2 Human toxicity non-cancer

Similarly to the observation made for the human toxicity cancer impact category, when using the EF 2.0 reference package, metals are the main contributors for human toxicity non-cancer (Table 55). If chromium does not contribute anymore, Zinc is now the dominant contributor for the large majority of representative products.

In EC-JRC-2018 non-weighted, although metals continue to play an important role, new inorganic and organic substances can be identified as significant contributors to the overall product impact score. JRC-2018 weighted provide a similar outcome with more elementary flows contributing to the 80% threshold.

Table 55: Elementary flows contributing up to 80% to total human toxicity non-cancer in all the 30 available EF reference products

CFs from:	Elementary flow	max relevance	relevant in datasets	total datasets analysed
EF 2.0	zinc emissions to agricultural soil	88%	28	30
	mercury emissions to air, unspecified	53%	16	30
	cadmium emissions to agricultural soil	41%	9	30
	arsenic v emissions to fresh water	25%	10	30
	zinc emissions to air, unspecified	28%	8	30
	zinc emissions to water, unspecified	18%	5	30
	zinc emissions to non-urban air or from high stacks	12%	8	30
	zinc emissions to fresh water	29%	2	30
	lead emissions to air, unspecified	13%	6	30
	zinc emissions to urban air close to ground	6%	4	30
	mercury emissions to urban air close to ground	8%	4	30
	lead emissions to non-urban air or from high stacks	10%	2	30
	arsenic emissions to non-urban air or from high stacks	6%	2	30
	cadmium emissions to non-urban air or from high stacks	6%	2	30
	lead emissions to agricultural soil	4%	1	30
EC-JRC-2018 Without robustness factor	carbon monoxide (fossil) emissions to air, unspecified	51%	24	30
	chloride emissions to fresh water	33%	24	30
	zinc emissions to agricultural soil	49%	18	30
	cadmium emissions to agricultural soil	69%	15	30
	chlorine emissions to urban air close to ground	54%	13	30
	lead emissions to agricultural soil	8%	11	30
	mercury emissions to air, unspecified	25%	10	30
	chlorine emissions to air, unspecified	32%	8	30
	chlorine emissions to fresh water	35%	7	30
	chlorine emissions to agricultural soil	25%	6	30
	chlorpyrifos emissions to air, unspecified	10%	2	30
	lead emissions to air, unspecified	8%	2	30
	dichlorvos emissions to air, unspecified	3%	1	30
	phorate emissions to agricultural soil	3%	1	30
	chromium vi emissions to air, unspecified	5%	1	30
	zinc emissions to fresh water	6%	1	30
	carbon monoxide (biogenic) emissions to urban air close to ground	6%	1	30
	fluoride emissions to fresh water	13%	1	30
EC-JRC-2018 with	mercury emissions to air, unspecified	60%	22	30
	carbon monoxide (fossil) emissions to air, unspecified	19%	20	30
	chloride emissions to fresh water	14%	18	30

CFs from:	Elementary flow	max relevance	relevant in datasets	total datasets analysed
robustness factor	lead emissions to agricultural soil	24%	18	30
	cadmium emissions to agricultural soil	93%	16	30
	zinc emissions to agricultural soil	17%	13	30
	chlorine emissions to urban air close to ground	30%	11	30
	mercury emissions to agricultural soil	9%	10	30
	lead emissions to air, unspecified	19%	9	30
	chlorpyrifos emissions to air, unspecified	21%	8	30
	lead emissions to non-urban air or from high stacks	13%	6	30
	mercury emissions to urban air close to ground	7%	6	30
	chlorine emissions to air, unspecified	17%	5	30
	morpholine emissions to water, unspecified	15%	4	30
	chlorine emissions to fresh water	17%	4	30
	dichlorvos emissions to air, unspecified	9%	3	30
	propargite emissions to air, unspecified	4%	3	30
	methane (biogenic) emissions to non-urban air or from high stacks	2%	3	30
	cadmium emissions to non-urban air or from high stacks	7%	2	30
	phorate emissions to agricultural soil	8%	2	30
	volatile organic compound emissions to air, unspecified	19%	2	30
	chlorine emissions to agricultural soil	7%	2	30
	arsenic emissions to non-urban air or from high stacks	7%	2	30
	carbon monoxide (biogenic) emissions to urban air close to ground	2%	1	30
	methane (biogenic) emissions to air, unspecified	10%	1	30

7.1.3 Ecotoxicity, freshwater

The contribution analysis performed on the freshwater ecotoxicity impact category provides a more balanced picture, with still metals contributing heavily in the majority of representative products (Zinc being again one the main contributor in 16 representative products) (Table 56).

The main difference between the EF 2.0 and the JRC-2018 CFs (non-weighted and weighted) are a reduced number of elementary flows contributing to the 80% threshold for the EC-JRC-2018 CFs and a more diverse composition of the type of substances contributing to the impact score.

Table 56: Elementary flows contributing up to 80% to total aquatic freshwater toxicity in all the 30 available EF reference products.

CFs from:	Elementary flow	max relevance	relevant in datasets	total datasets analysed
	zinc emissions to fresh water	84%	16	30
	folpet emissions to agricultural soil	55%	2	30
	morpholine emissions to water, unspecified	41%	4	30
	copper emissions to water, unspecified	40%	5	30
	copper emissions to agricultural soil	39%	16	30
	chromium emissions to fresh water	32%	15	30
	zinc emissions to water, unspecified	28%	8	30
EF 2.0	chlorpyrifos emissions to agricultural soil	25%	13	30
	cyfluthrin emissions to water, unspecified	23%	12	30
	chromium vi emissions to fresh water	19%	6	30
	prochloraz emissions to agricultural soil	18%	3	30
	vanadium emissions to air, unspecified	18%	4	30
	cypermethrin emissions to water, unspecified	17%	12	30
	chlorpyrifos emissions to water, unspecified	16%	12	30
	arsenic v emissions to fresh water	15%	6	30
	zinc emissions to air, unspecified	13%	2	30

CFs from:	Elementary flow	max relevance	relevant in datasets	total datasets analysed
	dimethyl sulphate emissions to air, unspecified	13%	1	30
	vanadium emissions to urban air close to ground	13%	1	30
	pyrene emissions to fresh water	11%	3	30
	copper emissions to non-urban air or from high stacks	11%	3	30
	vanadium emissions to fresh water	11%	2	30
	copper emissions to fresh water	11%	8	30
	nickel emissions to water, unspecified	11%	5	30
	zinc emissions to agricultural soil	10%	16	30
	chromium vi emissions to air, unspecified	10%	3	30
	dichlorvos emissions to agricultural soil	10%	4	30
	lambda-cyhalothrin emissions to water, unspecified	9%	12	30
	copper emissions to air, unspecified	9%	2	30
	antimony emissions to fresh water	9%	4	30
	cypermethrin emissions to agricultural soil	7%	12	30
	antimony emissions to air, unspecified	7%	2	30
	chlorothalonil emissions to agricultural soil	6%	1	30
	nickel emissions to fresh water	6%	2	30
	propanil emissions to agricultural soil	6%	5	30
	acetochlor emissions to agricultural soil	5%	1	30
	cyfluthrin emissions to air, unspecified	5%	10	30
	barium emissions to fresh water	5%	2	30
	chromium emissions to non-urban air or from high stacks	4%	3	30
	decane emissions to fresh water	4%	2	30
	isoproturon emissions to agricultural soil	4%	4	30
	lasso emissions to agricultural soil	4%	5	30
	chromium emissions to air, unspecified	4%	1	30
	bifenthrin emissions to water, unspecified	4%	4	30
	prochloraz emissions to water, unspecified	3%	1	30
	phorate emissions to agricultural soil	3%	4	30
	carbofuran emissions to agricultural soil	3%	8	30
	nitrobenzene emissions to fresh water	3%	1	30
	zinc emissions to non-urban air or from high stacks	3%	1	30
	cyfluthrin emissions to agricultural soil	3%	10	30
	antimony emissions to urban air close to ground	3%	1	30
	tebuconazole emissions to agricultural soil	3%	1	30
	phorate emissions to water, unspecified	3%	4	30
	aniline emissions to water, unspecified	3%	1	30
	terbuthylazin emissions to agricultural soil	2%	7	30
	fenvalerate emissions to water, unspecified	2%	2	30
	esfenvalerate emissions to water, unspecified	2%	4	30
	zinc emissions to urban air close to ground	2%	1	30
	cypermethrin emissions to air, unspecified	2%	5	30
	tannins emissions to water, unspecified	2%	1	30
	atrazine emissions to agricultural soil	2%	5	30
	carbendazim emissions to agricultural soil	2%	3	30
	aclonifen emissions to agricultural soil	2%	1	30
	chromium emissions to water, unspecified	1%	5	30
	fenvalerate emissions to agricultural soil	1%	1	30
	esfenvalerate emissions to agricultural soil	1%	3	30
	alpha-cypermethrin emissions to water, unspecified	1%	1	30
	simazine emissions to agricultural soil	1%	1	30
	metolachlor emissions to agricultural soil	1%	1	30
EC-JRC-2018 without robustness factor	chloride emissions to fresh water	83%	30	30
	sulfur emissions to fresh water	46%	18	30
	chlorpyrifos emissions to water, unspecified	45%	14	30
	phorate emissions to agricultural soil	38%	10	30
	hydrogen sulfide emissions to air, unspecified	35%	15	30

CFs from:	Elementary flow	max relevance	relevant in datasets	total datasets analysed
	sulfur emissions to agricultural soil	34%	2	30
	copper emissions to agricultural soil	21%	15	30
	calcium emissions to fresh water	9%	6	30
	lambda-cyhalothrin emissions to water, unspecified	8%	12	30
	phorate emissions to water, unspecified	8%	4	30
	zinc emissions to fresh water	7%	4	30
	bifenthrin emissions to water, unspecified	6%	4	30
	ammonia emissions to non-urban air or from high stacks	5%	10	30
	chlorpyrifos emissions to agricultural soil	5%	1	30
	esfenvalerate emissions to agricultural soil	5%	4	30
	ammonia emissions to air, unspecified	4%	8	30
	esfenvalerate emissions to water, unspecified	4%	3	30
	chlorpyrifos emissions to air, unspecified	3%	1	30
	carbofuran emissions to agricultural soil	3%	2	30
	bifenox emissions to water, unspecified	3%	2	30
	cyfluthrin emissions to water, unspecified	2%	1	30
	aluminium emissions to air, unspecified	2%	2	30
	deltamethrin emissions to water, unspecified	2%	1	30
	pirimiphos-methyl emissions to agricultural soil	2%	1	30
	bifenthrin emissions to air, unspecified	2%	2	30
	bifenox emissions to air, unspecified	2%	1	30
	chloride emissions to fresh water	68%	29	30
	chlorpyrifos emissions to water, unspecified	55%	15	30
	phorate emissions to agricultural soil	48%	10	30
	aluminium emissions to fresh water	45%	15	30
	sulfur emissions to agricultural soil	30%	2	30
	aluminium emissions to air, unspecified	28%	20	30
	sulfur emissions to fresh water	27%	11	30
	hydrogen sulfide emissions to air, unspecified	21%	5	30
	copper emissions to agricultural soil	18%	7	30
	aluminium emissions to agricultural soil	16%	7	30
	lambda-cyhalothrin emissions to water, unspecified	15%	15	30
	morpholine emissions to water, unspecified	14%	4	30
	bifenthrin emissions to water, unspecified	11%	4	30
	phorate emissions to water, unspecified	10%	8	30
	acetochlor emissions to agricultural soil	9%	6	30
	esfenvalerate emissions to agricultural soil	8%	10	30
EC-JRC-2018 with robustness factor	calcium emissions to fresh water	8%	3	30
	prochloraz emissions to agricultural soil	7%	1	30
	esfenvalerate emissions to water, unspecified	7%	9	30
	chlorpyrifos emissions to agricultural soil	6%	8	30
	bifenox emissions to water, unspecified	5%	2	30
	cyfluthrin emissions to water, unspecified	4%	6	30
	chlorpyrifos emissions to air, unspecified	4%	6	30
	carbofuran emissions to agricultural soil	4%	7	30
	deltamethrin emissions to water, unspecified	4%	2	30
	ammonia emissions to non-urban air or from high stacks	3%	2	30
	pirimiphos-methyl emissions to agricultural soil	3%	1	30
	bifenox emissions to air, unspecified	3%	1	30
	bifenthrin emissions to air, unspecified	3%	4	30
	cadmium emissions to agricultural soil	3%	5	30
	nitrobenzene emissions to fresh water	2%	1	30
	zinc emissions to fresh water	2%	2	30
	esfenvalerate emissions to air, unspecified	2%	3	30
	bifenthrin emissions to agricultural soil	2%	2	30
	cyfluthrin emissions to air, unspecified	2%	1	30

8 Normalisation

According to ISO 14044 (ISO, 2006b), normalisation in LCA is an optional step of Life Cycle Impact Assessment (LCIA). The normalisation factors represent the total impact of a reference region for a certain impact category (e.g. toxicity, etc.) in a reference year. For the EF, due to the international nature of supply chain, the use of global normalization factors are recommended.

The global normalisation factors (NF) reported in Table 57 are built on a vast collection of data on substance emissions into air, soil and water at global scale in 2010, as detailed in (Crenna et al., 2018).

Table 57. Global normalisation factors for toxicity related impact categories within the Environmental Footprint context.

Impact category	Unit	Global NF	Inventory coverage completeness[a]	Inventory robustness[b]
Freshwater ecotoxicity	CTUe	2.94E+14	III	III
Human toxicity cancer	CTUh	1.28E+05	III	III
Human toxicity non-cancer	CTUh	1.59E+06	III	III

(a) the extent to which the inventory data cover the list of flows available in ILCD, for each impact category: I=high (60% to 100%), II=medium (30% to 59%), III=low (0 to 29%)

(b) the quality of data, assessed by considering both the combination of different sources and the adoption of extrapolation strategies: I=high (data from published datasets from official data sources, subjected to a quality assurance procedure and limited use of extrapolation methods, i.e. <20 % of the impact derived from extrapolation), II=medium (non-publicly available or peer reviewed datasets and/or use of extrapolation methods for more than 20% but less than 80% of the impact), III=low (use of extrapolation methods for more than 80% of the impact)

Global normalisation factors derive from the characterization of 1585 elementary flows for ecotoxicity, 345 for human toxicity cancer and 1512 for human toxicity non-cancer.

The global inventory is built on the upscale of the EU inventory as available in (Sala et al., 2015), by using a factor 14.12 derived from (Cucurachi et al., 2014). This factor represents the ratio between the global extrapolated reference for mercury emissions and the related EU value reported by (Cucurachi et al., 2014).

For several substances, additional data from the current literature was retrieved and specific extrapolations adopted for refining or complementing the inventory, as follows.

Metals. Emissions to soil of metals proceeding from manure (arsenic, cadmium, chromium, copper, lead, mercury, nickel, zinc) were taken from (Leclerc & Laurent, 2017). Emissions to air of chromium, lead and antimony are based on the upscale of Chinese records (Cheng et al., 2014a; Tian et al., 2015), by means of the Chinese share of global electricity generated from coal (37%, (IEA, 2011)). Emissions to air of arsenic, cobalt, manganese and selenium were upscaled from Chinese records (Cheng et al., 2014b; Tian et al., 2015) to the global value by considering the Chinese share of global mercury emissions (31%, (UNEP, 2013)). Emissions of mercury to both air and water proceed from UNEP (2013). Emissions to water of aluminium come from Leclerc & Laurent (personal communication). Emissions to water of arsenic, cadmium, lead and selenium are based on the upscale of EU inventory from Sala et al. (2015) by considering the European share of global emissions to air, assumed to be the same as to water. Finally, emissions to water of cobalt, copper and manganese were come from the upscale of the updated EU inventory of (Leclerc & Laurent, 2017) by factor 14.12 derived from Cucurachi et al. (2014).

Pesticides. The EU inventory of pesticides in Sala et al. (2015) was replaced by its most up-to-date version from (Leclerc et al., 2019). It was complemented with emissions from three additional EU countries (Bulgaria, Croatia and Romania, previously neglected) and up-scaled based on the European share of global agricultural land (3.84%, (Faragò et al., 2019)).

Other emissions to air. For 58 substances, global emissions to air were retrieved from the available literature, as reported in Table 58.

Table 58. Global inventory of substances emitted to air, and related data sources differently from Sala et al. 2015.

CAS nr.	EF compliant substance name	Data source
431-89-0	1,1,1,2,3,3,3-heptafluoropropane*	(EC-JRC, 2013) (EDGAR, v.4.2 FT2010)
460-73-1	1,1,1,3,3-pentafluoropropane**	(EC-JRC, 2013) (EDGAR, v.4.2 FT2010)
71-55-6	1,1,1-trichloroethane	(Fraser et al., 2014)
76-13-1	1,1,2-trichlorotrifluoroethane	(Fahey & Hegglin, 2011)
1746-01-6	2,3,7,8-tetrachlorodibenzo-p-dioxin	(Fiedler et al., 2012)[2]
83-32-9	acenaphthene	(Shen et al., 2013b)
208-96-8	acenaphthylene	(Shen et al., 2013b)
74-86-2	acetylene	EC-EC-JRC 2018 (EDGAR v4.3.2_VOC_spec)
7664-41-7	ammonia	EC-EC-JRC 2016 (EDGAR v.4.3.1)
120-12-7	anthracene	(Shen et al., 2013b)
71-43-2	benzene	EC-EC-JRC 2018 (EDGAR v4.3.2_VOC_spec)
56-55-3	benzo[a]anthracene	(Shen et al., 2013b)
50-32-8	benzo[a]pyrene	(Shen et al., 2013b)
205-99-2	benzo[b]fluoranthene	(Shen et al., 2013b)
191-24-2	benzo[g,h,i]perylene	(Shen et al., 2013b)
207-08-9	benzo[k]fluoranthene	(Shen et al., 2013b)
630-08-0	carbon monoxide	(EC-JRC, 2013) (EDGAR, v.4.2 FT2010)
75-69-4	CFC-11	(Fraser et al., 2014)
75-71-8	CFC-12	(Fahey & Hegglin, 2011)
74-87-3	chloromethane	(EC-JRC, 2018a)(EDGAR v4.3.2_VOC_spec)
218-01-9	chrysene	(Shen et al., 2013b)
74-84-0	ethane	(EC-JRC, 2018a)(EDGAR v4.3.2_VOC_spec)
74-85-1	ethylene	(EC-JRC, 2018a)(EDGAR v4.3.2_VOC_spec)
206-44-0	fluoranthene	(Shen et al., 2013b)
86-73-7	fluorene	(Shen et al., 2013b)
74-83-9	halon-1001***	Leclerc & Laurent (personal communication)
1717-00-6	HCFC-141b	(Fraser et al., 2014)
75-68-3	HCFC-142b	(Fraser et al., 2014)
118-74-1	hexachlorobenzene	Leclerc & Laurent (personal communication)
110-54-3	hexane	(EC-JRC, 2018a)(EDGAR v4.3.2_VOC_spec)
354-33-6	HFC-125	(EC-JRC, 2013) (EDGAR, v.4.2 FT2010)
811-97-2	HFC-134a	(EC-JRC, 2013) (EDGAR, v.4.2 FT2010)
75-37-6	HFC-152a	(EC-JRC, 2013) (EDGAR, v.4.2 FT2010)
75-46-7	HFC-23	(EC-JRC, 2013) (EDGAR, v.4.2 FT2010)
690-39-1	HFC-236fa	EC-EC-JRC 2013 (EDGAR, v.4.2 FT2010)
75-10-5	HFC-32	EC-EC-JRC 2013 (EDGAR, v.4.2 FT2010)
138495-42-8	HFC-43-10-mee	EC-EC-JRC 2013 (EDGAR, v.4.2 FT2010)
78-79-5	isoprene	EC-EC-JRC 2018 (EDGAR v4.3.2_VOC_spec)
74-82-8	methane (biogenic)	EC-EC-JRC 2013 (EDGAR, v.4.2 FT2010)
91-20-3	naphthalene	Shen et al. 2013
106-97-8	n-butane	EC-EC-JRC 2018 (EDGAR v4.3.2_VOC_spec)
10024-97-2	nitrous oxide	EC-EC-JRC 2013 (EDGAR, v.4.2 FT2010)
109-66-0	pentane	EC-EC-JRC 2018 (EDGAR v4.3.2_VOC_spec)
85-01-8	phenanthrene	Shen et al. 2013
74-98-6	propane	EC-EC-JRC 2018 (EDGAR v4.3.2_VOC_spec)
115-07-1	propene	EC-EC-JRC 2018 (EDGAR v4.3.2_VOC_spec)
129-00-0	pyrene	Shen et al. 2013
2551-62-4	sulfur hexafluoride	EC-EC-JRC 2013 (EDGAR, v.4.2 FT2010)
108-88-3	toluene	EC-EC-JRC 2018 (EDGAR v4.3.2_VOC_spec)
25551-13-7	trimethylbenzene	EC-EC-JRC 2018 (EDGAR v4.3.2_VOC_spec)

1) Average emission per person as reported in (Fiedler et al., 2012), upscaled to global population in NF underpinning inventory as: * HFC-227ea; **HFC-245fa; *** bromomethane.

---

[2] Average emission per person as reported in Fiedler et al. 2012, upscaled to global population

94

After its classification into the EF compliant elementary flows, the final inventory was characterized by using the characterization factors (CF) developed in the Environmental Footprint context and presented in chapter 6. Regarding the specificity of the emission compartment, "emission to air (soil or water), unspecified" CFs were used. For those substances and groups for which there was no possibility to be mapped into an existing EF compliant elementary flow, ad hoc CFs were calculated in order to improve the coverage (Table 59). Furthermore, acknowledging the potential underestimation of the NF especially for freshwater ecotoxicity due to a limited list of substances, the unmapped pesticides were characterizing by means of a proxy CF derived as average of the available CF for pesticides in the normalization inventory. Details are in the online supplementary material.

Uncertainties in the calculation of the global normalisation factors may derive from different sources, namely the reliability of data sources, the mapping of elementary flows, and the extrapolations from EU to global emissions.

Table 59: Substances and groups available in the inventory of normalisation, for which a specific CF was calculated.

Substance/group name, as in the inventory of global NF	Possible mapping into EF compliant elementary flow	Rules for calculation CF when a direct mapping was not possible
Adsorbable organic halogens (AOX)	Adsorbable organic halogen compounds*	
Beta-cyfluthrin	Cyflutrin	
Brominated diphenylethers (PBDE)	Decabromophenyl ether	
BTEX	-	Characterized by using 50th %ile of CF of benzene, m-diethylbenzene, ethylene, m-xylene, o-diethylbenzene, o-xylene, p-xylene, toluene, xylene (all isomers).
Chlorides (as total Cl)	Chlorides, unspecified*	
Copper chelate	Copper metal (conservative approach)	
Copper salt	Copper metal (conservative approach)	
Cyanides (as total CN)	Cyanide*	
Fluorides (as total F)	Hydrogen fluoride*	
Halogenated organic compounds (as AOX)	Adsorbable organic halogen compounds*	
Octylphenols and Octylphenol ethoxylates	4-(1,1,3,3-tetramethylbutyl)phenol	
Other fungicides	Fungicides, unspecified*	
Other herbicides	Herbicides, unspecified*	
Other insecticides	Insecticides, unspecified*	
Petroleum oils	Oils, unspecified*	
Phenols (as total C)	Phenol	
Polycyclic aromatic hydrocarbons (PAH)	Polycyclic aromatic hydrocarbons*	
Tetracopper-tricalciumsulfate	Copper metal (conservative approach)	
Trichlorobenzenes (TCBs) (all isomers)	-	Characterized by using 50th %ile of 1,2,3-trichlorobenzene; 1,2,4-trichlorobenzene; 1,3,5-trichlorobenzene; trichlorobenzene
Zeta-cypermethrin	Cypermethrin	

*: See chapter 6.1.4

9 Conclusions

The present work was performed in order to provide new characterisation factors for the freshwater ecotoxicity, human toxicity cancer and human toxicity non-cancer impact categories for the EU Environmental Footprint.

Those CFs were calculated using new physicochemical properties and toxicity data extracted from the REACH-IUCLID database of the European Chemical Agency (ECHA), from the OpenFoodTox of the European Food Security Authorities (EFSA), and from the Pesticide Properties database (PPDB) of the University of Hertforshire.

All substance characterization factors were calculated using the USEtox® 2.1 model.

Although great care was put in selecting the input data, the use of an automated extraction procedure applied on the REACH-IUCLID database (in total more than 6 million individual cells) is not error free. Therefore, any substances contributing at an exceptional very high level to a product toxicity score should be scrutinized. Although the underlying data are not directly available from the online supplementary information, due to property and confidentiality reasons, all those data are available on the ECHA dissemination website (https://echa.europa.eu/home).

Similarly, all the data extracted from OpenFoodTox and from the 'PPDB' database are freely accessible on their respective web site: (https://www.efsa.europa.eu/en/microstrategy/openfoodtox and https://sitem.herts.ac.uk/aeru/ppdb/en/atoz.htm respectively).

With the new CFs, the EC-JRC has achieved major progress in relation to different aspects:

— The input data (physicochemical and toxicity properties) have been improved using more consistent and robust sources like ECHA and EFSA database.

— The coverage of elementary flows used in EF has been significantly broaden (6011 CFs for freshwater ecotoxicity compared to 2499 with USEtox® 2.1; 3450 new CFs human toxicity cancer compared to 426 with USEtox® 2.1; for human toxicity cancer the number has not changed: 621).

— The newly introduced robustness factor on CFs level reflects the capability of the underlying multimedia fate model in terms of adequately characterising different groups of substances (organics, inorganics, metal non essentials, metal essentials)

All the EC-JRC-2018 CFs are to be used in the context of the EU Environmental Footprint, and the level of recommendation is III.

Since the outcomes of the Pellston workshop (UNEP- SETAC, 2018) require a significant update of the USEtox® 2.1 model for both freshwater ecotoxicity, human toxicity cancer, and humna toxicity non-cancer, the EC will decide if and how to take into account the new version of USEtox® including new CFs, once they become available.

References

Azimonti, G., Galimberti, F., Marchetto, F., Menaballi, L., Ullucci, S., Pellicioli, F., …van der Voet, H. (2015). Comparison of NOEC values to EC10/EC20 values, including confidence intervals, in aquatic and terrestrial ecotoxicological risk assessment. *EFSA Supporting Publications*, *12*(12), 906E. https://doi.org/doi:10.1126/scisignal.2004112

Barbaro, B., Baldin, R., Kovarich, S., Pavan, M., Fioravanzo, E., & Bassan, A. (2015). Further development and update of EFSA's Chemical Hazards Database. *EFSA Supporting Publications*, *12*(7). https://doi.org/10.2903/sp.efsa.2015.EN-823

Bare, J. (2011). TRACI 2.0: the tool for the reduction and assessment of chemical and other environmental impacts 2.0. *Clean Technologies and Environmental Policy*, *13*(5), 687–696. https://doi.org/10.1007/s10098-010-0338-9

Beasley, A., Belanger, S. E., Brill, J. L., & Otter, R. R. (2015). Evaluation and comparison of the relationship between NOEC and EC10 or EC20 values in chronic Daphnia toxicity testing. *Environmental Toxicology and Chemistry*, *34*(10), 2378–2384. https://doi.org/10.1002/etc.3086

Boethling, R. S., Howard, P. H., Meylan, W., Stiteler, W., Beauman, J., & Tirado, N. (1994). Group contribution method for predicting probability and rate of aerobic biodegradation. *Environmental Science & Technology*, *28*(3), 459–465. https://doi.org/10.1021/es00052a018

Castellani, V., Sala, S., & Saouter, E. G. (2018). Analysis of the contribution of elementary flows to toxicity-related impact categories in the Consumer Footprint indicator. Presentation at: *SETAC case study symposium,* 24-26 September 2018. Wien.

Cheng, H., Zhou, T., Li, Q., Lu, L., & Lin, C. (2014a). Anthropogenic chromium emissions in China from 1990 to 2009. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0087753

Cheng, K., Tian, H. Z., Zhao, D., Lu, L., Wang, Y., Chen, J., … Huang, Z. (2014b). Atmospheric emission inventory of cadmium from anthropogenic sources. *International Journal of Environmental Science and Technology*, *11*(3), 605–616.

Crenna, E., Secchi, M., Benini, L., & Sala, S. (2018). Global environmental impacts: data sources and methodological choices for calculating normalisation factors for LCA. *The International Journal of Life Cycle Assessment*. https://doi.org/doi.org/10.1007/s11367-018-1534-9

Cucurachi, S., Sala, S., Laurent, A., & Heijungs, R. (2014). Building and characterizing regional and global emission inventories of toxic pollutants. *Environmental Science & Technology*, *48*(10), 5674–5682.

Diamond, M. L., Gandhi, N., Adams, W. J., Atherton, J., Bhavsar, S. P., Bulle, C., … Vijver, M. G. (2010). The clearwater consensus: the estimation of metal hazard in fresh water. *The International Journal of Life Cycle Assessment*, *15*(2), 143–147. https://doi.org/10.1007/s11367-009-0140-2

Dorne, J. Lou, Richardson, J., Kass, G., Georgiadis, N., Monguidi, M., Pasinato, L., … Robinson, T. (2017). Editorial: OpenFoodTox: EFSA's open source toxicological database on chemical hazards in food and feed. *EFSA Journal*, *15*(1), 2016–2018. https://doi.org/10.2903/J.EFSA.2017.E15011

Dourson, M. L., Andersen, M. E., Erdreich, L. S., & MacGregor, J. A. (2001). Using human data to protect the public's health. *Regulatory Toxicology and Pharmacology*, *33*(2), 234–256. https://doi.org/10.1006/rtph.2001.1469

EC-JRC, (European Commission - Joint Research Center). (2003). Technical Guidance Document on Risk Assessment. *European Chemicals Bureau*, *Part II*, 7–179.

EC-JRC, (European Commission - Joint Research Center). (2011a). *ILCD Handbook: Recommendations for Life Cycle Impact Assessment in the European context*. https://doi.org/10.278/33030

EC-JRC, (European Commission - Joint Research Center). (2011b). *Recommendations based on existing environmental impact assessment models and factors for life cycle assessment in European context. Luxembourg: Publications Office of the European Union. EUR24571EN. ISBN 978-92-79- 17451-3.*

EC-JRC, (European Commission - Joint Research Center). (2013). Emission Database for Global Atmospheric Research (EDGAR), released version 4.2 FT2010. Retrieved from http://edgar.jrc.ec.europa.eu/overview.php?v=42FT2010, accessed December 2018

EC-JRC, (European Commission - Joint Research Center). (2018a). Global speciated NMVOC Emissions: EDGAR v4.3.2_VOC_spec (January 2017). Retrieved November 15, 2018, from http://edgar.jrc.ec.europa.eu/overview.php?v=432_VOC_spec accessed December 2018

EC-JRC, (European Commission - Joint Research Center). (2018b). EF Reference Package 3.0. Retrieved from http://eplca.jrc.ec.europa.eu/LCDN/developerEF.xhtml accessed December 2018

EC, (European Commission). (2006). Regulation (EC) No 1907/2006 of the European Parliament and the council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), *L396*, 1–394.

EC, (European Commission). (2013a). Commission Recommendation (2013/179/EU) of 9 April 2013 on the use

of common methods to measure and communicate the life cycle environmental performance of products and organisations. *Official Journal of the European Communities*, *56*(L124). https://doi.org/10.3000/19770677.L_2013.124.eng

EC, (European Commission). (2013b). Communication from the commission to the European Parliament and the council. Building the Single Market for Green Products. COM(2013) 196.

ECB, (European Chemical Bureau). (2003). European Commision - Technical Guidance Document on Risk Assessment - Part 1. *Part II*, 337. https://doi.org/10.1002/mp.12308

ECHA, (European Chemical Agency). (2014). *Guidance on information requirements and chemical safety assessment Chapter R11 : PBT/vPvB assessment*. Helsinky, Finland.

ECHA, (European Chemical Agency). (2015). *Guidance on Information Requirements and Chemical Safety Assessment Chapter R .7a : Endpoint specific guidance*. Retrieved from https://echa.europa.eu/documents/10162/13632/information_requirements_r7a_en.pdf accessed December 2018

ECHA, (European Chemical Agency). (2017). *Guidance on information requirements and chemical safety assessment Chapter R.7b: Endpoint specific guidance*. Helsinky, Finland. Retrieved from https://echa.europa.eu/documents/10162/13632/information_requirements_r7b_en.pdf accessed December 2018

ECHA, (European Chemical Agency). (2018a). IUCLID 6. Retrieved from https://iuclid6.echa.europa.eu/ accessed December 2018

ECHA, (European Chemical Agency). (2018b). Registration statistics infograph - ECHA. Retrieved from https://echa.europa.eu/registration-statistics-infograph# accessed December 2018

Fahey, D. W., & Hegglin, M. I. (2011). *Twenty Questions and Answers about the Ozone Layer 2010 Update: Scientific Assessment of Ozone Depletion 2010*. World Meteorological Organisation.

Fantke, P., Bijster, M., Hauschild, M. Z., Huijbregts, M., Jolliet, O., Kounina, A., … Van Zelm, R. (2017). *USEtox 2.0 Documentation (Version 1.00)*. https://doi.org/10.11581/DTU:00000011

Faragò, M., Benini, L., Sala, S., Secchi, M., & Laurent, A. (2019). National inventories of land occupation and transformation flows in the world for land use impact assessment. *International Journal of Life Cycle Assessment*, *submitted*.

Fiedler, H., Cao, Z., Huang, J., Wang, B., Deng, S., & Yu, G. (2012). PCDD/PCDF Inventories 1990 vs. 2012. *Organohalogen Compound*, *74*, 1521–1524.

Fraser, P. J., Dunse, B. L., Manning, A. J., Walsh, S., Wang, R. H. J., Krummel, P. B., … O'Doherty, S. (2014). Australian carbon tetrachloride emissions in a global context. *Environmental Chemistry*, *11*(1), 77–88.

Frischknecht, R., & Jolliet, O. (2016). *Global Guidance for Life Cycle Impact Assessment Indicators. Vol 1*. Paris, France. Retrieved from https://www.lifecycleinitiative.org/training-resources/global-guidance-lcia-indicators-v-1/ accessed December 2018

Hauschild, M. Z., Huijbregts, M., Jolliet, O., Macleod, M., Margni, M., Van De Meent, D., … McKone, T. E. (2008). Building a model based on scientific consensus for life cycle impact assessment of chemicals: The search for harmony and parsimony. *Environmental Science & Technology, 42(19),7032-7037* https://doi.org/10.1021/es703145t

Hauschild, M. Z., & Potting, J. (2005). Spatial Differentiation in Life Cycle Impact Assessment - The EDIP2003 methodology. *Environmental News*, *80*, 1–195. Retrieved from https://www2.mst.dk/udgiv/publications/2005/87-7614-579-4/pdf/87-7614-580-8.pdf

Henderson, A. D., Hauschild, M. Z., van de Meent, D., Huijbregts, M. a. J., Larsen, H. F., Margni, M., … Jolliet, O. (2011, June). USEtox fate and ecotoxicity factors for comparative assessment of toxic emissions in life cycle analysis: sensitivity to key chemical properties. *The International Journal of LCA, 16, 701-709* https://doi.org/10.1007/s11367-011-0294-6

Huijbregts, M. A., Guinée, J. B., & Reijnders, L. (2001). Priority assessment of toxic substances in life cycle assessment. III: Export of potential impact over time and space. *Chemosphere*, *44*(1), 59–65.

IEA, (International Energy Agency). (2011). *Statistics and Balances*. Retrieved from https://www.iea.org/ accessed December 2018

ISO, (International Standard Organisation). (2006a). *ISO 14040. Environmental management -- Life cycle assessment -- Principles and framework*. Geneva, Switzerland.

ISO, (International Standard Organisation). (2006b). ISO 14044. *Environmental Management- Life Cycle Assessment-Requirements and Guidelines*. Geneva, Switzerland

Jolliet, O., Margni, M., Charles, R., Humbert, S., Payet, J., & Rebitzer, G. (2003). Presenting a New Method IMPACT 2002 + : A New Life Cycle Impact Assessment Methodology, *The International Journal of LCA, 8*(6), 324–330.

Kemna, R., Van Elburg, M., & Holsteijn, V. (2005). *MEEuP methodology report, final*.

Klimisch, H. J., Andreae, M., & Tillmann, U. (1997). A systematic approach for evaluating the quality of experimental toxicological and ecotoxicological data. *Regulatory Toxicology and Pharmacology*, *25*(1), 1–5. https://doi.org/10.1006/rtph.1996.1076

Leclerc, A., & Laurent, A. (2017). Framework for estimating toxic releases from the application of manure on agricultural soil: national release inventories for heavy metals in 2000–2014. *Science of The Total Environment*, *590*, 452–460.

Leclerc, A., Sala, S., Secchi, M., & Laurent, A. (2019). Building national emission inventories of toxic pollutants in Europe. *submitted to Environment International*.

Lewis, K. A., Tzilivakis, J., Warner, D. J., & Green, A. (2016). An international database for pesticide risk assessments and management. *Human and Ecological Risk Assessment: An International Journal*, *22*(4), 1050–1064. https://doi.org/10.1080/10807039.2015.1133242

Lyman, W. J., Reehl, W. F., & Rosenblatt, D. H. (1990). Handbook of Chemical Property Estimation Methods. Environmental Behaviour of Organic Compounds. American Chemical Society: Washington, DC, Vol. 960

McKone, T. E., & Enoch, K. G. (2002). CalTOX, A multimedia total exposure model spreadsheet user's guide. Version 4.0. *Lawrence Berkeley National Laboratory*. https://doi.org/10.2172/803756

Meylan, W., & Howard, P. (1995). Atom/fragment contribution method for estimating octanol–water partition coefficients. *Journal of Pharmaceutical Sciences*, *84*(1), 83–92. https://doi.org/10.1002/jps.2600840120

Meylan, W., Howard, P. H., & Boethling, R. S. (1992). Molecular topology/fragment contribution method for predicting soil sorption coefficients. *Environmental Science & Technology*, *26*(8), 1560–1567. https://doi.org/10.1021/es00032a011

Meylan, W. M., & Howard, P. H. (1991). Bond contribution method for estimating henry's law constants. *Environmental Toxicology and Chemistry*, *10*(10), 1283–1293. https://doi.org/10.1002/etc.5620101007

Meylan, W. M., Howard, P. H., & Boethling, R. S. (1996). Improved method for estimating water solubility from octanol/water partition coefficient. *Environmental Toxicology and Chemistry*, *15*(2), 100–106. https://doi.org/10.1002/etc.5620150205

Müller, N., de Zwart, D., Hauschild, M., Kijko, G., & Fantke, P. (2017). Exploring REACH as a potential data source for characterizing ecotoxicity in life cycle assessment. *Environmental Toxicology and Chemistry*, *36*(2), 492–500. https://doi.org/10.1002/etc.3542

Neely, W. B. (1985). *Environmental Exposure From Chemicals: Volume I.* CRC Press: Boca Raton, p. 254. Retrieved from https://www.crcpress.com/Environmental-Exposure-From-Chemicals-Volume-I/Neely/p/book/9781315892689

OECD, (Organisation for Economic Cooperation and Development). (1981). Decision of the Council concerning the Mutual Acceptance of Data in the Assessment of Chemicals. *Acts OECD*, *30*(81), 1–2. Retrieved from http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=263&InstrumentPID=339&Lang=en&Book=False accessed December 2018

OECD, (Organisation for Economic Cooperation and Development). (1998). *OECD Guideline for the testing of chemicals. Repeated Dose 90-day Oral Toxicity Study in Non-Rodents*. Retrieved from https://www.oecd-ilibrary.org/docserver/9789264070721-en.pdf?expires=1542279363&id=id&accname=ocid194935&checksum=5D61471D3EE4C301BCD38C9F5A0C9454 accessed december 2018

OECD, (Organisation for Economic Cooperation and Development). (2015). Test No. 422: Combined Repeated Dose Toxicity Study with the Reproduction/Developmental Toxicity Screening Test. p. 14. https://doi.org/10.1787/9789264242692-en

OECD, (Organisation for Economic Cooperation and Development). (2018). The OECD QSAR Toolbox. Available at http://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm accessed December 2018

Pepelko, W. E., & Withey, J. R. (1985). Methods for Route-To-Route Extrapolation of Dose. *Toxicology and Industrial Health*, *1*(4), 153–175. https://doi.org/10.1177/074823378500100410

Posthuma, L., Suter, G. W., & Traas, T. P. (2002). *Species sensitivity distributions in ecotoxicology*. CRC Press: Boca Raton, p.617.

PPDB (Pesticide Properties Data Base). (2017). *The University of Hertfordshire Agricultural Substances Database Background and Support Information*. Retrieved from https://sitem.herts.ac.uk/aeru/ppdb/en/docs/Background_and_Support.pdf Accessed December 2018

Rosenbaum, R. K. (2015). Ecotoxicity. In M. Z. Hauschild & M. A. J. Huijbregts (Eds.), *Life Cycle Impact Assessment* (LCA Compendium, pp. 139–162). Springer Netherlands. https://doi.org/10.1007/978-94-017-9744-3_8

Rosenbaum, R. K., Bachmann, T. M., Gold, L. S., Huijbregts, M. A. J., Jolliet, O., Juraske, R., … Hauschild, M. Z. (2008). USEtox - The UNEP-SETAC toxicity model: Recommended characterisation factors for human toxicity and freshwater ecotoxicity in life cycle impact assessment. *International Journal of Life Cycle Assessment*, *13*(7), 532–546. https://doi.org/10.1007/s11367-008-0038-4

Sala, S., Benini, L., Beylot, A., Castellani, V., Cerutti, A., Corrado, S., … Pant, R. (2018). *Consumption and Consumer Footprint: methodology and results. Indicators and Assessment of the environmental impact of EU consumption.* Luxembourg. https://doi.org/10.2760/15899

Sala, S., Benini, L., Mancini, L., & Pant, R. (2015). Integrated assessment of environmental impact of Europe in 2010: data sources and extrapolation strategies for calculating normalisation factors. *The International Journal of Life Cycle Assessment*, *20*(11), 1568–1585.

Saouter, E. G., Aschberger, K., Fantke, P., Hauschild, M. Z., Bopp, S. K., Kienzler, A., … Sala, S. (2017a). Improving substance information in USEtox®, part 1: Discussion on data and approaches for estimating freshwater ecotoxicity effect factors. *Environmental Toxicology and Chemistry*, *36*(12), 3450–3462. https://doi.org/10.1002/etc.3889

Saouter, E. G., Aschberger, K., Fantke, P., Hauschild, M. Z., Kienzler, A., Paini, A., … Sala, S. (2017b). Improving substance information in USEtox®, part 2: Data for estimating fate and ecosystem exposure factors. *Environmental Toxicology and Chemistry*, *36*(12), 3463–3470. https://doi.org/10.1002/etc.3903

Saouter, E. G., De Schryver, A., Pant, R., & Sala, S. (2018). Estimating chemical ecotoxicity in EU ecolabel and in EU product environmental footprint. *Environment International,* 118, 44-47. https://doi.org/10.1016/j.envint.2018.05.022

Schröder, K., Escher, S. E., Hoffmann-Dörr, S., Kühne, R., Simetska, N., & Mangelsdorf, I. (2016). Evaluation of route-to-route extrapolation factors based on assessment of repeated dose toxicity studies compiled in the database RepDose®. *Toxicology Letters*, *261*, 32–40. https://doi.org/10.1016/j.toxlet.2016.08.013

Shen, H., Huang, Y., Wang, R., Zhu, D., Li, W., Shen, G., … Tao, S. (2013a). Global atmospheric emissions of polycyclic aromatic hydrocarbons from 1960 to 2008 and future predictions. *Environmental Science & Technology*, *47*(12), 6415–6424. https://doi.org/10.1021/es400857z

Simulations-Plus. (2016). Admet predictor™. Lancaster, California 93534-7059. Retrieved from https://www.simulations-plus.com/software/admetpredictor/

Tian, H. Z., Zhu, C. Y., Gao, J. J., Cheng, K., Hao, J. M., Wang, K., … Zhou, J. R. (2015). Quantitative assessment of atmospheric emissions of toxic heavy metals from anthropogenic sources in China: historical trend, spatial distribution, uncertainties, and control policies. *Atmospheric Chemistry and Physics*, *15*(17), 10127–10147.

UNEP, (United Nations Environmental Program). (2013). *Global Mercury Assessment 2013: Sources, Emissions, Releases, and Environmental Transport*. *UNEP report,* (Vol. 42). https://doi.org/DTI/1636/GE

US-EPA, (United States Environmental Protection Agency). (2012). On-Line BIOWIN™ User's Guide (v4.10).

US-EPA, (United States Environmental Protection Agency). (2018). Estimation Programs Interface Suite™ for Microsoft® Windows, v 4.11. Washington, DC, USA. Retrieved from https://www.epa.gov/tsca-screening-tools/epi-suitetm-estimation-program-interface#what accessed December 2018

UNEP -SETAC, (Society of Environmental Toxicology and Chemistry). (2018). Pellston and Technical Workshops. Retrieved from https://www.setac.org/general/custom.asp?page=PTWorkshops accessed December 2018

Van Zelm, R., Huijbregts, M. A. J., & Van De Meent, D. (2009). USES-LCA 2.0-a global nested multi-media fate, exposure, and effects model. The *International Journal of Life Cycle Assessment*, *14*(3), 282–284. https://doi.org/10.1007/s11367-009-0066-8

Wegmann, F., Cavin, L., MacLeod, M., Scheringer, M., & Hungerbühler, K. (2009). The OECD software tool for screening chemicals for persistence and long-range transport potential. *Environmental Modelling & Software*, *24*(2), 228–237. https://doi.org/10.1016/j.envsoft.2008.06.014

Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Modeling*, *28*(1), 31–36. https://doi.org/10.1021/ci00057a005

List of abbreviations and definitions

ACR	Acute-Chronic Ratio
AD	Applicability Domain
ADI	Acceptable Daily Intake
Agr.soilC	Continental agricultural soil compartment
AirC	Continental air compartment
AirU	Urban air compartment
ANS	food Additives and Nutrient Sources added to food
AOEL	Acceptable Operator Exposure Limit
ARfD	Acute Reference Dose
ASTM	American Society for Testing and Materials
BAF	Bio-Accumulation Factor
BCF	Bio-Concentration Factor
BMDL	BenchMark Dose Lower bound
bw	body weight
CAS	Substance Abstract Service
CDV	Critical Dilution Volume
CEF	food Contact material, Flavourings and processing aids
ChV	Chronic Value
CLP	Classification, Labelling and Packaging
CF	Characterization Factor
CnF	Conversion Factor
CMR	Carcinogenic, Mutagenic or toxic to Reproduction
CONTAM	Contaminants in the food chain
CPDB	Carcinogenic Potency DataBase
CSA	Chemical Safety Assessment
CTU	Comparative Toxic Unit
CTUe	Comparative Toxic Unit, ecotoxicity
CTUh	Comparative Toxic Unit, human health
DOC	Dissolved Organic Carbon
EC	European Commission
ECHA	European Substance Agency
EC-JRC	European Commission - Joint Research Centre
EC50	median Effect Concentration
ECb50	median Effect Concentration on biomass/yield (equivalent to ECy50)
ECr50	median Effect Concentration on reproduction
ED50	median Effective Dose
EDGAR	Emission Database for Global Atmospheric Research
EF	Environmental Footprint
EfF	Effect Factor
ExF	Extrapolation Factor
EFSA	European Food Safety Authority
EPA	united states Environmental Protection Agency
ERA	Environmental Risk Assessment

ESR Endpoint Study record

EUSES European Union System for the Evaluation of Substances

FDA united states Food and Drugs Administration

FEEDAP Additives and Products or substances used in animal FEED

FETAX Frog Embryo Teratogenesis Assay in Xenopus

FF Fate Factor

Fr.waterC Continental Fresh water compartment

GHG Green-House Gases

GLP Good Laboratory Practices

GM Geometric Mean

GCV% Geometric Coefficient of Variation (expressed in percentage)

HBGV Health-Based Guidance Value

HC50 median Hazardous Concentration

HH Human health

HPLC High Performance Liquid Chromatography

IC50 median Immobilisation Concentration (equivalent to Inhibition Concentration)

ILCD International reference Life Cycle Data system

IRIS DB Integrated Risk Information System DataBase

IS Impact score

ITIS Integrated Taxonomy Information System

IUCLID International Uniform Chemical Information Database

IUCN International Union for the Conservation of Nature

Kdoc Dissolved (colloidal) organic carbon/Water partitioning coefficient

Koc Organic carbon/Water partitioning coefficient

Kow n-Octanol/Water partitioning coefficient (equivalent of *pow*)

KpSED Sediment/Water partitioning coefficient

KpSOIL Soil/Water partitioning coefficient

KpSS Suspended solids/water partitioning coefficient (equivalent of Ksus)

LC50 median Lethal Concentration

LCA Life Cycle Assessment

LCDN Life Cycle Data Network

LCIA Life Cycle Impact Assessment

LD50 median Lethal Dose

LL50 median Lethal Level

LOAEC Lowest Observed Adverse Effect Concentration

LOAEL Lowest Observed Adverse Effect Level

LOEC Lowest Observed Effect Concentration

LOEL Lowest Observed Effect Level

MATC Maximum Acceptable Toxicant Concentration

MCI Molecular Connectivity Index

MTDI Maximum Tolerable Daily Intake

MW Molecular Weight

Nat.soilC Continental natural soil compartment

NAEL No Adverse Effect Level

NDA   Dietetic products, Nutrition and Allergies

NEL   No Effect Level

NF   Normalisation Factor

NMVOC   Non-Methane Volatile Organic Carbons

NOAEC   No Observed Adverse Effect Concentration

NOAEL   No Observed Effect Level

NOEC   No Observed Effect Concentration

NOEL   No Observed Effect Level

NOELR   No observed Effect Loading Rate

NOTEL   No Observed Toxic Effect Level

OECD   Organisation for Economic Co-operation and Development

OEF   Organisation Environmental Footprint

PAF   Potentially Affect Fraction

PAH   Polycyclic Aromatic Hydrocarbons

PBT   Persistent, Bio-accumulative and Toxic

PCB   PolyChlorinated Biphenyls

PEC   Predicted Environmental Concentration

PEF   Product environmental footprint

PNEC   Predicted No Effect Concentration

PRAPeR   Plant Protection Products and their Residues

QSAR   Quantitative Structure Activity Relationship

REACH   Registration, Evaluation, Authorisation and restriction of Substances

RDT   Repeated Dose Toxicity

seawaterC Continental sea water compartment

SETAC   Society of Environmental Toxicology And Chemistry

SMILES   Simplified Molecular Input Line Entry System

SR   Study Report

SSD   Species Sensitivity Distribution

SVHC   Substances of Very High Concern

TAB   Technical Advisory Board

TD50   median Toxic Dose

TDI   Tolerable Daily Intake

TGD   Technical Guidance Document

TLm   median Tolerance Limit

UUID   Universal Unique IDentifier

UNEP   United Nations Environmental Programme

UVCB   Unknown or Variable Composition

VOC   Volatile Organic Carbons

vPvB   very Persistent, very Bio-accumulative

WAF   Water Accommodated Fraction

WHO   World Health Organisation

WoE   Weight of Evidence

WWTP   Waste-Water Treatment Plant

XF   Exposure Factor

Definitions

Applicability domain (AD): The Applicability Domain (AD) of a QSAR model is that part of the multi-dimensional substance space where the model has been developed, and for which predictions for new compound can be considered reliable.

CAS number: Substance Abstracts Service maintains the most comprehensive list of substance substances. Each substance registered in the CAS Registry is assigned a CAS Registry Number. The CAS Registry Number is widely used as a identifier of substance substances. Further Information: http://www.cas.org.

EC number: number allocated by the Commission of the European Communities as a term used to replace the EINECS / ELINCS / NLP number designation. This number is a seven-digit system, separated into 3 groups by hyphens of the type XXX-XXX-X. EC numbers starts by 2 or 3 for substances belonging to EINECS (Existing Substances), 4 for ELINCS (New Substances) and 5 for NLP (No-Longer Polymers).

Endpoint: An endpoint is an observable or measurable inherent property of a substance. It can for example refer to a physical property like vapour pressure or degradability or to a biological effect that a given substance has on human health or the environment, e.g. carcinogenicity, irritation, aquatic toxicity. A toxic endpoint is the result of a study conducted to determine how dangerous a substance is. The data collected from such studies are used to report the relative toxicity of the compound to various regulatory agencies and environmental compliance groups. Toxic endpoints can include mortality, behaviour, reproductive status or physiological changes.

Good Laboratory Practices (GLP): Good Laboratory Practice (GLP) is a quality system concerned with the organisational process and the conditions under which non-clinical health and environmental safety studies are planned, performed, monitored, recorded, archived and reported.

Klimisch scores: It represents a scoring system to assess the reliability of data; particularly from toxicological and ecotoxicological studies, that may be extended to physicochemical and environmental fate and behaviour studies. The Klimisch scoring system is based on 4 categories, i.e. Klimisch 1 (reliable without restrictions), 2 (reliable with restrictions), 3 (not reliable), 4 (not assignable).

Mono-constituent substance: As a general rule, a substance, defined by its composition, in which one main constituent is present to at least 80% (w/w).

Plant Protection Products/Substances: Active substances and preparations containing one or more active substances, put up in the form in which they are supplied to the user, intended to: i) protect plants or plant products against all harmful organisms or prevent the action of such organisms, ii) influence the life processes of plants, other than as a nutrient, (e.g. growth regulators), iii) preserve plant products, iv) destroy undesired plants or v) prevent undesired growth of plants.

Quantitative structure-Activity Relationship (QSAR): It is the relationship between the physical and\or substance properties of a substance and their ability to cause a particular effect. The goal of QSAR studies in toxicology is to develop whereby the toxicity of a substance can be predicted from its substance structure by analogy with the properties of other toxic substances of known structure and toxic properties. In practice QSARs are mathematical models used to predict the properties of substances from their molecular structure.

Read-across: Read-across is a technique for predicting endpoint information for one substance (target substance), by using data from the same endpoint from one or more source substances, which are considered to be similar.

Reference point: Defined point on an experimental dose–response relationship for the critical effect. This term is synonymous to point of departure (USA). Reference points include the lowest or no observed adverse effect level (LOAEL/NOAEL) or

benchmark dose lower confidence limit (BDML), used to derive a reference value or Margin of Exposure in human and animal health risk assessment. In the ecological area, these include lethal dose (LD50), effect concentration (EC5/ECx), no (Adverse) effect concentration/dose (NOEC/NOAEC/NOAED), no (adverse) effect level (NEL/NOAEL), hazard concentration (HC5/HCx) derived from a Species Sensitivity Distributions (SSD) for the ecosystem.

Reference value: The estimated maximum dose (on a body mass basis) or the concentration of an agent to which an individual may be exposed over a specified period without appreciable risk. Reference values are derived by applying an uncertainty factor to the reference point. Examples of reference values in human health include acceptable daily intake (ADI) for food and feed additives, pesticides and food contact materials, tolerable upper intake levels (UL) for vitamins and minerals, and tolerable daily intake (TDI) for contaminants. For acute effects and operators, the acute reference dose (ARfD) and the acceptable operator exposure level (AOEL).In animal health and the ecological area, these include maximum tolerated dose (MTD) and predicted no effect concentration (PNEC) respectively.

List of figures

List of tables

Annexes: List of supplementary materials available online.

All supplementary materials mentioned in this report are available at:
http://eplca.jrc.ec.europa.eu/LCDN/developerEF.xhtml.

- Physicochemical properties tables
  - **Kow.** List of variables available for n-octanol/water partition coefficient.
  - **Koc.** List of variables available for adsorption/desorption partition coefficients (Koc, KpSED, KpSUSP and KpSOIL).
  - **Vap25.** List of variables available for vapour pressure.
  - **Sol25.** List of variables available for water solubility.
  - **HENRY.** List of variables available for Henry's Law constant.
  - **Degradation rate.** List of variables available for degradation screening test and half life.
- Freshwater ecotoxicity tables
  - **Taxonomy DB.** Database of species with taxonomy information.
  - **Species as in REACH-IUCLID.** Number and percentage of species in REACH-IUCLID database.
  - **Reference point per SSD group.** Number and percentage of reference point in REACH-IUCLID.
  - **Test duration per SSD group.** List of test duration recorded in REACH-IUCLID.
  - **Report SSD.** Number of species available for each SSD group for each chemical.
  - **Algae, Crustacean, Fish**. Number of reference points per the three main trophic levels, divided on acute and chronic tests.
- Human toxicity tables
  - **HH variables.** List of variables available for human toxicity, ingestion and inhalation exposure routes.
  - **Conversion ingestion.** Formulas and parameters used to convert ingestion exposure repeated dose toxicity values to mg/kg bw/day.
  - **Conversion inhalation.** Formulas and parameters used to convert inhalation exposure repeated dose toxicity values to mg/m3.
  - **Reference points ingestion.** List and relative frequency of all reference points included in the oral toxicity REACH-IUCLID database.
  - **Reference points inhalation.** List and relative frequency of all reference points included in the inhalation toxicity REACH-IUCLID database.
  - **Guidelines ingestion.** List of guidelines and their associated USEtox® endpoint categories. Ingestion exposure.
  - **Guidelines inhalation.** List of guidelines and their associated USEtox® endpoint categories. Inhalation exposure.
  - **Case studies.** Case studies: Repeated dose via oral toxicity – 16 discrepancies found in the sample of 250 model results, with related explanation.
- R codes
  - **R code physchem.** R script used to derive physico chemical properties from the REACH-IUCLID database.
  - **R code humantox.** R script used to derive human toxicity non cancer ED50 from the REACH database.
  - **R code ecotox.** R script used to derive ecotoxicity HC20 from the REACH-IUCLID database.
  - **R code PPDB.** R script used to derive information from PPDB. For physico chemical properties, freshwater ecotoxicity and human non cancer toxicity.
  - **R additional codes.** R scripts used to 1) generate the new input table to be run in USEtox®; 2) correct metals from REACH-IUCLID, apply weights and harmonise USEtox® and JRC impact categories; 3) generate CFs for proxies.
- New input table for USEtox® model.
- Characterization factors (as outcome of USEtox® 2.1 model).
- Characterization factors (for ILCD emission compartments).
- Contribution analysis.
- Normalization factors.
- List of uncharacterized flows.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications
You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

**The European Commission's
science and knowledge service**

Joint Research Centre

JRC Mission

As the science and knowledge service
of the European Commission, the Joint
Research Centre's mission is to support
EU policies with independent evidence
throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub – Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office
of the European Union