Journal of Toxicological Studies 2025, 3(1), 1798.
https://doi.org/10.59400/jts1798

Article

The evolution of the EFSA OpenFoodTox database

Nadia Iovine[1,*], Alessandra Roncaglioni[1], Luca Sartori[2], Chihae Yang[3], Emilio Benfenati[1]

[1] Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Department of Environmental Health Science, Via Mario Negro, 2,20156 Milano, Italy
[2] S-IN Soluzioni Informatiche, Via Giuseppe Ferrari, 14,36100 Vicenza, Italy
[3] Molecular Networks GmbH, Neumeyerstraße, 24,90411 Nürnberg, Germany
*** Corresponding author:** Nadia Iovine, nadia.iovine@marionegri.it

CITATION

Iovine N, Roncaglioni A, Sartori L, et al. The evolution of the EFSA OpenFoodTox database. Journal of Toxicological Studies. 2025; 3(1): 1798.
https://doi.org/10.59400/jts1798

ARTICLE INFO

Received: 13 November 2024
Accepted: 25 December 2024
Available online: 3 January 2025

Abstract: Since its establishment in 2002, the European Food Safety Authority (EFSA) has been providing independent scientific advice on risks associated with the food chain. This manuscript provides a description of EFSA's chemical hazards database OpenFoodTox (OFT), future perspectives and activities. OFT aims at mapping all the hazard identification and characterisation data that have been published in outputs from EFSA throughout the years. To date, OFT contains data for more than 5700 chemical substances in the food/feed chain. In line with the One Substance-One Assessment approach as part of the Chemicals Strategy for Sustainability, EFSA aims to further improve data quality and interoperability of OFT with IUCLID 6 and the EU Common Data Platform on Chemical Safety. To enhance its usability as a supporting tool for risk assessment activities, OFT will be migrated to IUCLID 6. More data will be collected and added to OFT, including endpoints related to in vitro assays, non-critical effects and exposure values. Furthermore, new *in silico* models (e.g., tools for read-across and grouping) will be developed based on the data already present in OFT for chemicals and endpoints that have been tested, with the aim of estimating the corresponding properties for the untested chemicals and endpoints.

Keywords: hazard assessment; risk assessment; toxicology; ecotoxicology; OFT; IUCLID; *in silico* model

1. Introduction

The European Food Safety Authority (EFSA) is committed to protecting the European citizens with particular attention to adverse effects associated to food and feed. Established in 2002, EFSA has performed thousands of risk assessment evaluations for the substances that may have a critical impact. These evaluations require the work of key experts in their specific fields, who continuously contribute to advancing the knowledge about the impact of substances. The substances evaluated by EFSA belongs to different classes, such as food ingredients, food additives and flavourings, vitamins, novel foods, food contact materials, feed additives, toxins, plant protection products, veterinary substances and contaminants. The properties of interest are those related to human health risk assessment (HHRA), ecological risk assessment (ERA) and animal health risk assessment (AHRA). Plant protection and plant health are also within EFSA's remit. This is quite peculiar, since other sister agencies are also involved in HHRA and ERA, but not so much in AHRA and plant health. From the broad series of chemical substances to the diversity of the risk assessments, EFSA deals with multiple properties for multiple organisms, not only humans. The overall evaluation of such a complex body of properties and substances is a challenge and for this reason, EFSA also investigates how modern technologies can help to fully exploit

the large amount of data available and to address the existing data gaps. Several activities have been promoted in this direction: for instance, EFSA published a Guidance on the Weight-of-evidence to combine multiple lines of evidence, often arising from heterogeneous sources [1]. This perspective, which is a methodological one, is highly beneficial to define the roadmap for the use of data produced with new approach methods (NAMs), thus allowing to use more sources of data as inputs for substance assessment.

The huge amount of work, which was done by many experts to gather and organize data, has generated many opinions and curated property values. To organise the data on tens of endpoints for thousands of substances, EFSA created and made available the OpenFoodTox (OFT) database [2]. This initiative is consistent with the European Union and EFSA's policy [3] to "Widen EFSA's evidence base and optimise access to its data", with the aim of promoting open access to data relevant for safety. EFSA has critically evaluated and made use of the past work done by the experts within the panels to build up a very sound source of information.

OFT is publicly available, and has been integrated with other platforms to increase the reuse of data. For example, OFT is currently integrated with the AMBIT database[1], U.S. EPA CompTox Chemicals DashBoard[2], and U.S. EPA Toxicity Estimation Software Tool (TEST)[3].

The data collected in OFT has proved to be very useful in developing *in silico* models [2–7], due to the fact that (1) the quality of the data is very high and that (2) OFT is a unique collection of data on endpoints, which offered the possibility to develop models covering a wider range of ecological targets, beyond the classical models for the three trophic levels used for aquatic toxicity: algae, daphnia and fish.

EFSA is devoted to continuously improve OFT by populating it with more substances and data, and by expanding the set of available properties of the substances (e.g., by including physicochemical and toxicokinetic properties).

This report aims to inform about the advancements of OFT, newly introduced properties, and properties which are planned to be added in the next years, along with the integration of data which will be derived from the implementation and use of NAMs. The decision to include these new data expands the scope of OFT, which was originally created to collect hazard characterization data. Through the addition of toxicokinetics data (and the associated physicochemical properties) OFT will become part of a larger platform dedicated to risk assessment.

2. The OpenFoodTox database structure

OFT summarizes data for hazard identification and characterization of human, animals and environmental health. Since its creation, OFT has been enriched and updated with data from EFSA documents (opinions, statements, and conclusions) on risk assessment of food and feed.

Reported below are the four main areas of focus of the database:

- Chemical identification: the entity that has been assessed in the EFSA opinions or statements is described in terms of nomenclature (e.g., EU nomenclature, CAS number, IUPAC name), chemical formula, and structure (e.g., SMILES).

- Document: this database section contains the document descriptors of the EFSA opinion or statement from which the data have been extracted and stored in the database (e.g., title of the document, DOI, name of the EFSA Panel).
- Hazard identification: this section of the database reports information regarding the genotoxicity/carcinogenicity status of the assessed substance and the critical study which was used to derive the health-based guidance value (Tolerable daily intake - TDI, Acceptable daily intake - ADI), or the margin of exposure values or the margin of safety values. More specifically, the database hosts toxicity data on human health, animal health (target and non-target species), and ecotoxicity (soil and water compartments).
- Hazard characterisation/risk characterisation: this section provides information on the health-based guidance value (hazard characterisation), margin of exposure or the margin of safety (risk characterisation) and environmental standards (hazard characterisation or risk characterisation).

Moreover, in 2023 the database has been updated with physicochemical properties, absorption, distribution, metabolism, and excretion (ADME) values and toxicokinetic data from EFSA documents and other sources too. Data is collected and reported following a data model based on the Organisation for Economic Co-operation and Development (OECD) Harmonised Templates for Reporting Chemical Test Summaries (OHTs).

Users can freely download the full OFT database (the latest version was released on 13 September 2023) from the EFSA Knowledge Juction community on Zenodo[4]; alternatively, a specific dataset/datasheet can be downloaded using the OFT MicroStrategy Dashboard[5].

On the Zenodo website, it is possible to download several spreadsheets:

- OpenFoodToxTX22809_2023.xlsx: reporting the full database;
- substance characterisation: reporting the chemical information and characterization of each substance present in the database (e.g., CAS Number, EC number, Molecular formula, SMILES notation);
- physicochemical properties: reporting data collected for physicochemical properties and toxicokinetic studies;
- EFSA outputs: reporting information on all the EFSA outputs used for the collection of data (e.g., title, date of publication, doi);
- reference points: reporting 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 risk (e.g., ADI, Threshold of toxicological concern (TTC), Acute reference dose (ARfD), Acceptable operator exposure level (AOEL));
- reference values: reporting the defined point on an experimental dose–response relationship for the critical effect, e.g.,Lowest/No observed adverse effect level (LOAEL/NOAEL), Benchmark dose level (BMDL), No observed effect concentration (NOEC), No observed adverse effect concentration (NOAEC), No observed adverse effect dose (NOAED),, Effective concentration (ECx), Lethal dose 50 (LD50);
- genotoxicity: reporting data on genotoxicity studies.

OFT covers the work of many units and panels (**Figure 1**), including ANS (Food Additives and Nutrient Sources Added to Food), CEF (Food Contact Materials, Enzymes, Flavourings and Processing Aids), CONTAM (Contaminants in the Food Chain), FEEDAP (Additives and Products or Substances used in Animal Feed), NDA (Dietetic Products, Nutrition and Allergies), PPR panel and PRAPeR unit (Plant Protection Products and their Residues).

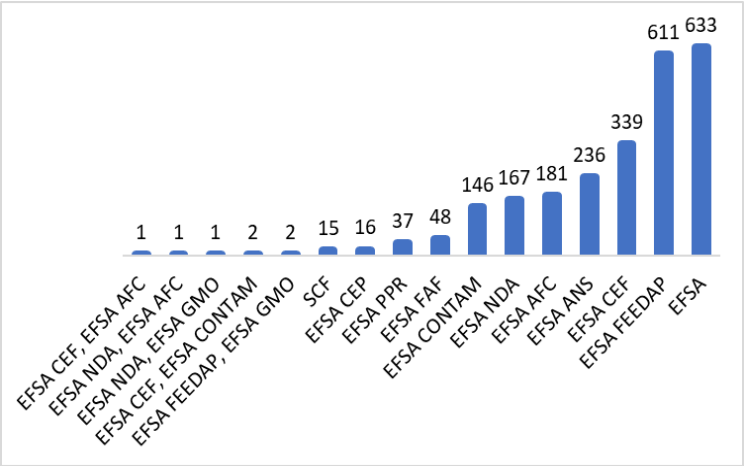

Figure 1. EFSA outputs contained in OFT grouped by their author(s); old studies do not report the detail of the specific Panel or Unit, thus they are simply indicated as EFSA.

To date, OFT includes documents that were published from 2000 to August 2022. A total of 10,827 assessments were produced for 5251 substances that are currently registered in OFT, by analysing a total of 2437 EFSA documents (**Figure 2**). The highest number of documents regards pesticides. Food-related substances are also broadly represented. Even if some categories, such as zootechnical additives, may seem to be poorly represented, OFT represents a relevant source of data for substances that are poorly studied. The number of substances within the different categories is analogous to the number of available assessments. Pesticides are the most represented category, followed by flavourings. The No category group refers to the Novel Food opinion.

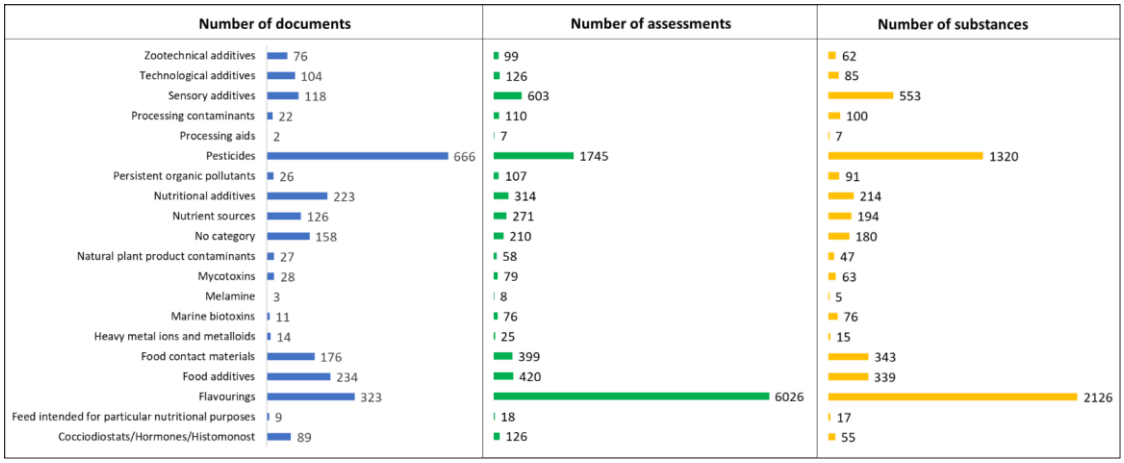

Figure 2. number of documents (blue), assessments (green) and substances (yellow) available for each category of compounds present in OFT version 6 (13 September 2023).

As illustrated in **Figure 3**, OFT stores a large variety of information, ranging from data related to human health to animal health (target and non-target species) and ecotoxicological health (water and soil compartment) data. Ecotox (water compartment) is the most represented category, as OFT contains a large number of data related to the water compartment, and a similar amount of data for the soil compartment. Furthermore, data on human health and animal non-target species health are also very abundant. The high diversity of ecological and animal species is very important and it is necessary to properly protect all the species. Unfortunately, in the common practice, most of the studies address very few aquatic species, while most of the other species are completely neglected. OFT contributes to cover this gap of knowledge.

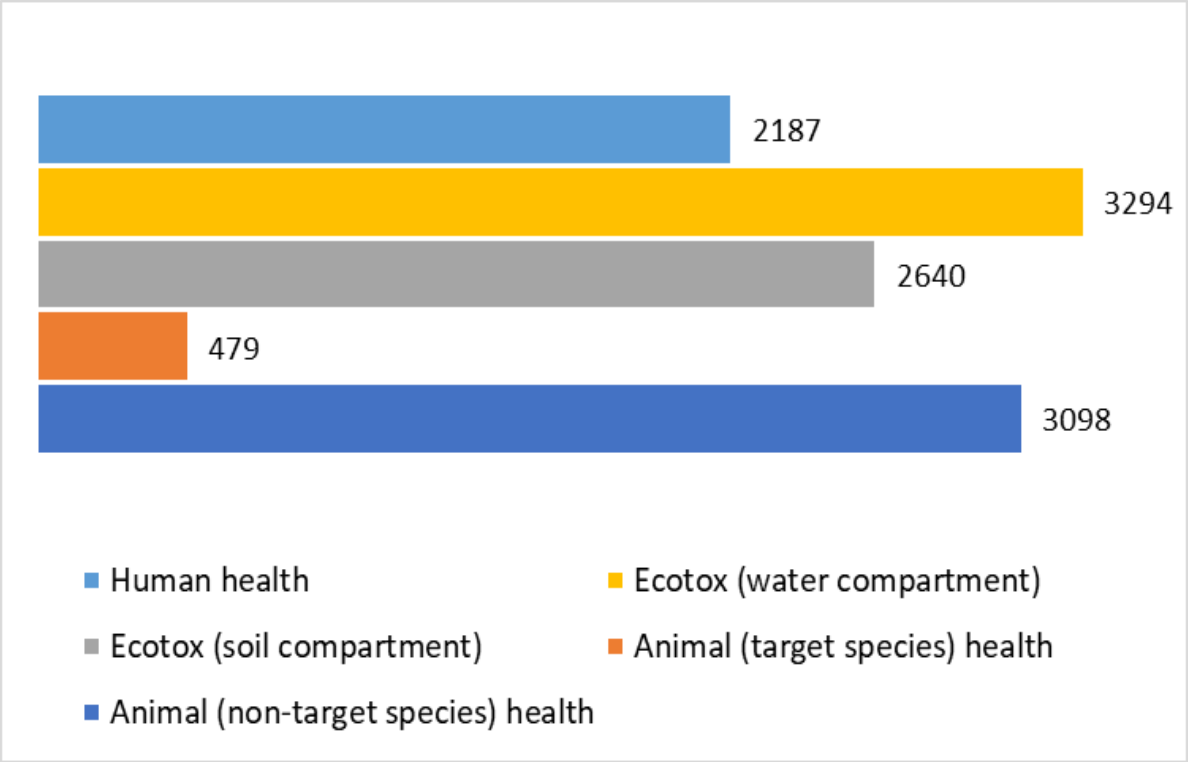

Figure 3. Summary of the data present in OFT for human, ecotoxicological and animal health.

Studies for 41 endpoints (**Table 1**) are reported in OFT. There are thousands of data available for the following endpoints: LD50, NOAEL, NOEC, EC50, and LC50. Thus, acute toxicity is well represented, but there are also many data available for subchronic effects.

Table 1. The number of data available within OFT for the different endpoints and parameters.

Endpoint/Parameter	number of data	Endpoint	number of data
LD50 (Half maximum Lethal Dose)	2289	IC50 (Half maximum Inhibition Concentration)	5
NOAEL (No Observed Adverse Effect Level)	2259	EAC (Estimated Acceptable Concentration)	3
NOEC (No Observed Effect Concentration)	2042	ED50 (Effective Dose 50)	3

Table 1. (*Continued*).

Endpoint/Parameter	number of data	Endpoint	number of data
EC50 (Half maximum Effective Concentration)	1501	EL50 (Effective Loading Rate resulting in 50% effect)	3
LC50 (Half maximum Lethal Concentration)	1419	LC10 (Lethal Concentration at which 10% effect is observed	3
dose level	470	BMD05 (Benchmark Dose Level)	2
LR50 (internal Lethal Residues)	421	EC15 (Effect Concentration at which 15% effect is observed)	2
NOEL (No Observed Effect Level)	412	EC25 (Effect Concentration at which 25% effect is observed)	2
ER50 (50% Effective Rates)	316	EC5 (Effect Concentration at which 5% effect is observed)	2
LOAEL (Lowest Observed Adverse Effect Level)	122	NOAR (No Effect Application Rate)	2
TEF (Toxic Equivalency Factor)	75	EC20 (Effect Concentration at which 20% is observed)	1
EC10 (Effect Concentration at which 10% effect is observed)	73	LD100 (Absolute Lethal Dose)	1
BMDL10 (Benchmark Dose Level 10)	63	LDLo (Lethal Dose Low)	1
RPF (Relative Potency Factor)	55	Lethal potency	1
BMDL05 (Benchmark Dose Level 5)	39	LOEC (Lowest Observed Effect Concentration)	1
conc. Level	34	MLD (Minimum Lethal Dose)	1
NOAEC (No Observed Adverse Effect Concentration)	28	NOAEDD (No Observed Adverse Effect Dietary Dose)	1
LDD50 (Lethal Dietary Dose)	21	NOEDD (No Observed Effect Dietary Dose)	1
LOEL (Lowest Observed Effect Level)	9	T5 (Tumour dose at which 5 % of increase of tumours is observed)	1
BMDL01(Benchmark Dose Level 1)	7	Tumour Incidence 25% Increase	1
BMDL (Benchmark Dose Level)	6	Total	11,698

The study reported in the database has, if available, the information on the type of study (short-term toxicity; acute toxicity; subchronic; chronic/long term toxicity; reproduction toxicity; study with volunteers), along with the organism/cells used. To date, the studies collected include data on 153 organisms that belong to different species.

In the human health area, chronic and subchronic studies are the most abundant, followed by reproductive and acute studies (**Figure 4**). In some cases, the study type is ‘not reported’; this means that the information on the study type (type of toxicity study) is not provided in the opinion. Although the study type is not reported, the endpoint and the associated value are provided in OFT.

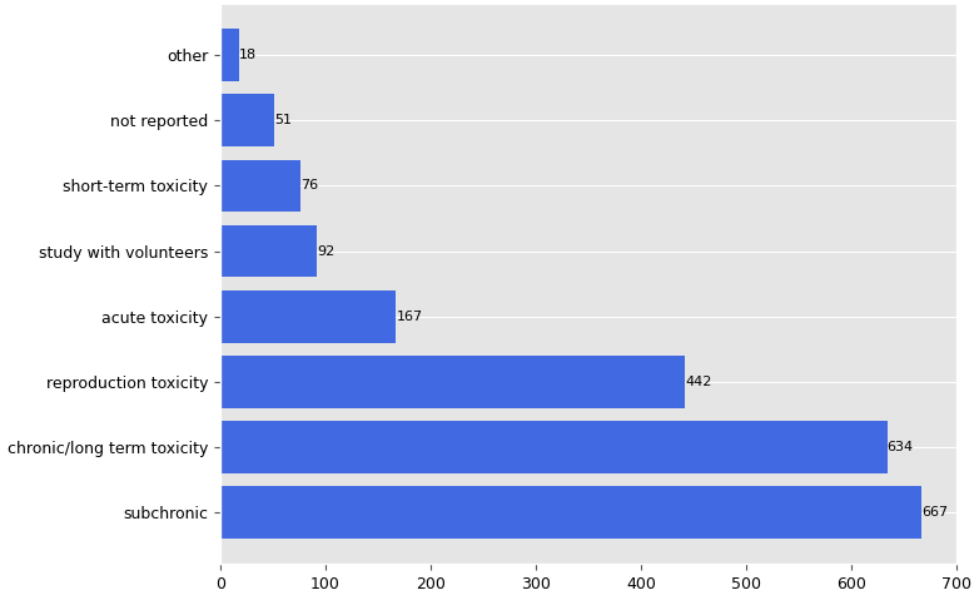

Figure 4. Toxicity studies for the human health area - Number of studies by each 'TESTTYPE'.

The number of toxicity and epidemiological studies in the database for each species related to the human health hazard assessment are reported in **Figure 5**. Most of the data derives from studies on rat, followed by dog, mouse, rabbit, human and pig, while for 63 data the information on the species is missing

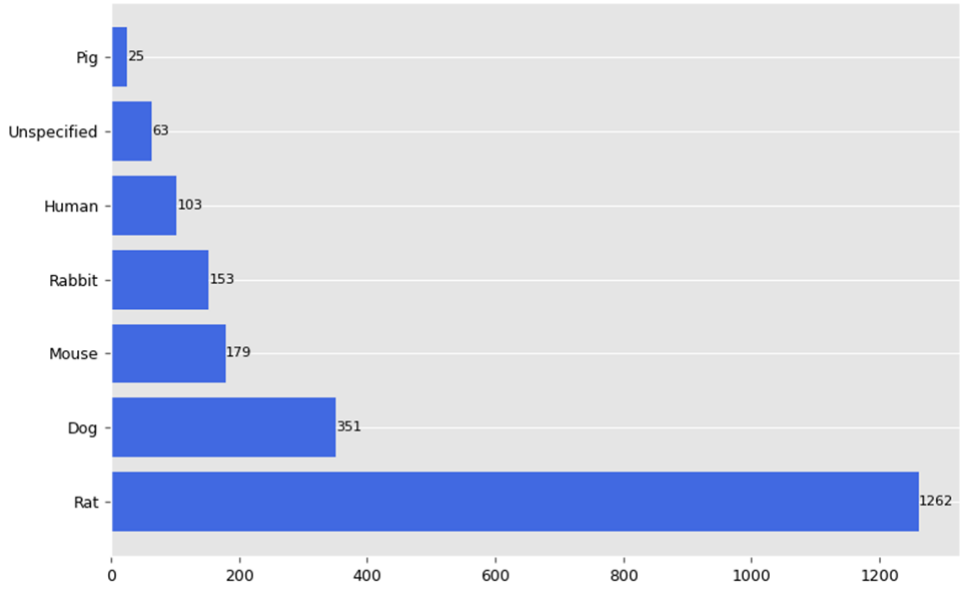

Figure 5. Classification of toxicity and epidemiological studies in OpenFoodTox for the human health area. Data entries in the database are shown for sub-chronic and chronic studies for reported species with sample size > 15.

In the animal health area, the majority of the data come from acute studies followed by reproduction and short-term toxicity studies (**Figure 6**).

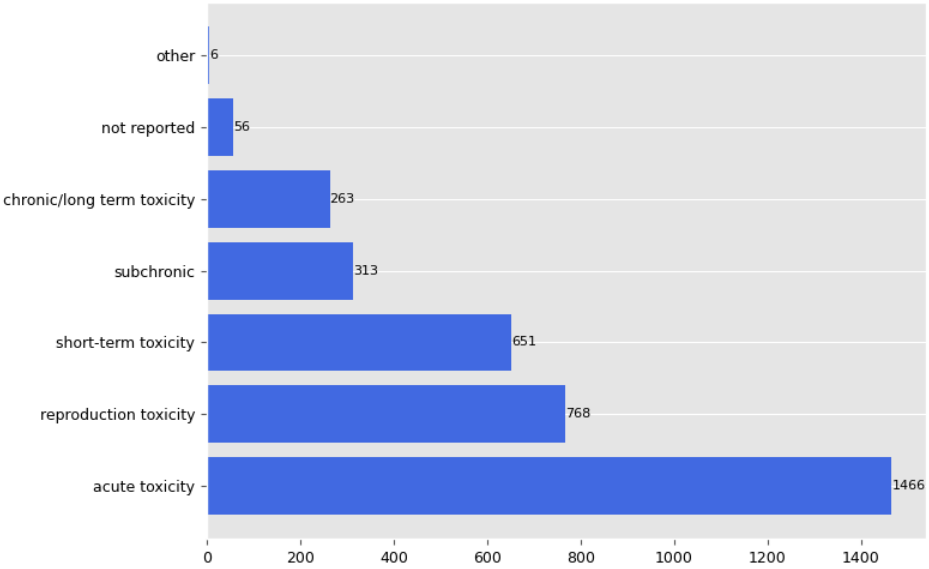

Figure 6. Toxicity studies for the animal health area - Number of studies by each 'TESTTYPE'.

Figure 7 shows the number of toxicity studies present in the database for each species related to the animal health hazard assessment. Most of the data derive from studies on rat, mallard duck and bobwhite quail.

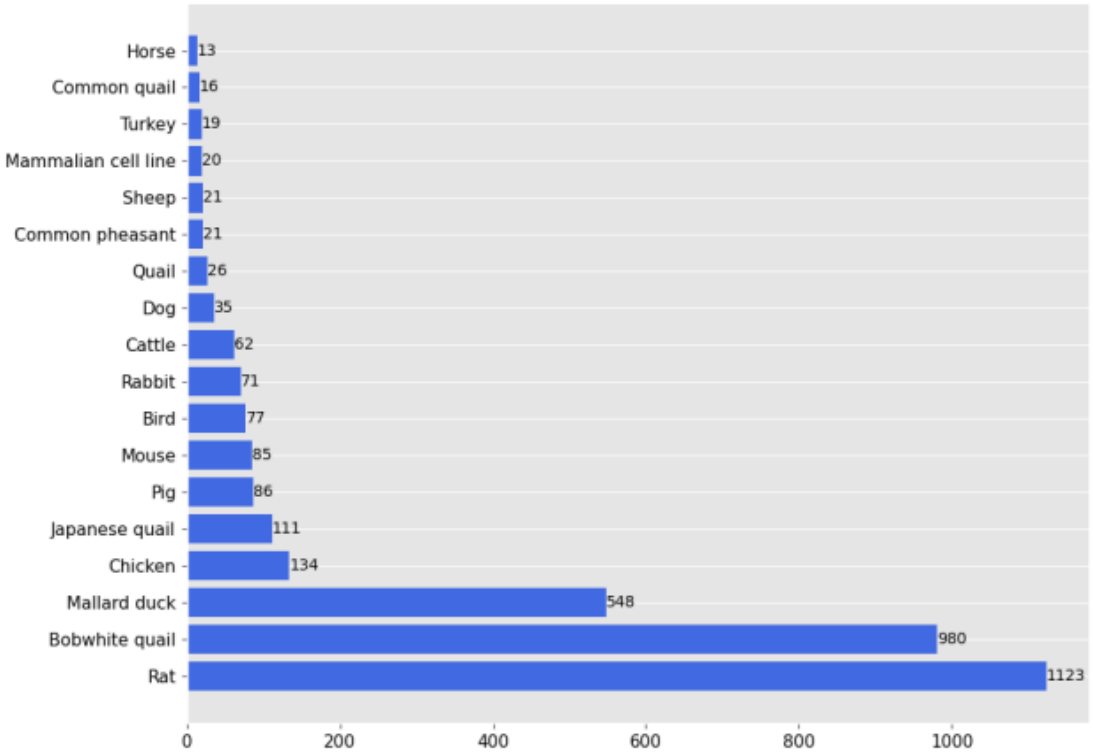

Figure 7. Classification of toxicity studies in OpenFoodTox for the animal health area. Data entries in the database are shown for sub-chronic and chronic studies for reported species with sample size > 10.

The OpenfoodTox database is also a repository for studies related to ecotoxicology. The number of studies reported in the database for each species related to ecotox (water and soil compartment) are reported in **Figure 8** (acute studies) and **Figure 9** (chronic/long term studies).

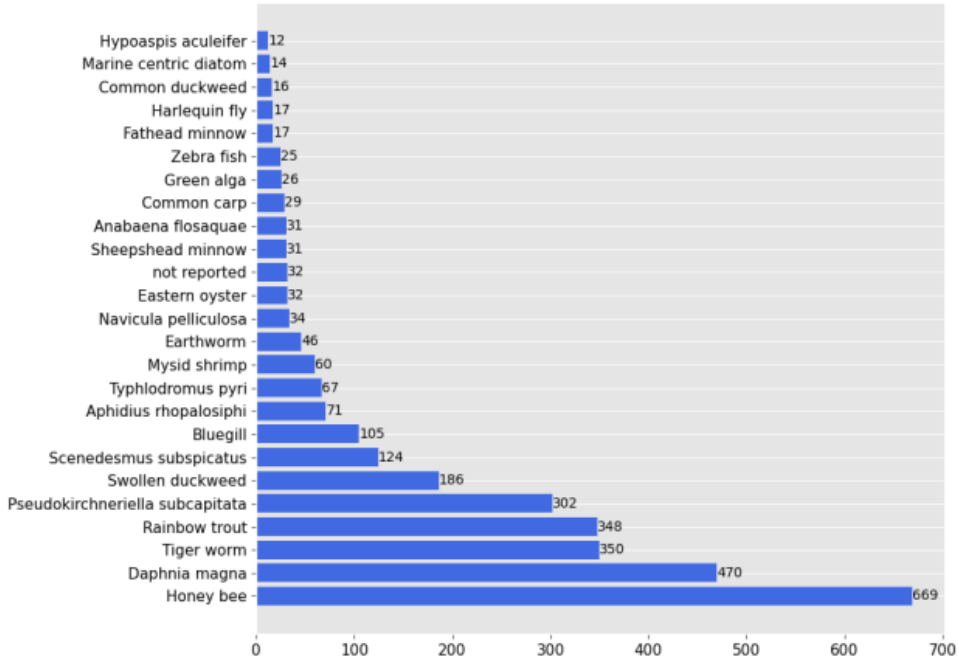

Figure 8. Classification of toxicity studies in OpenFoodTox for ecotox. Data entries in the database are shown for acute toxicity studies for reported species with sample size > 10.

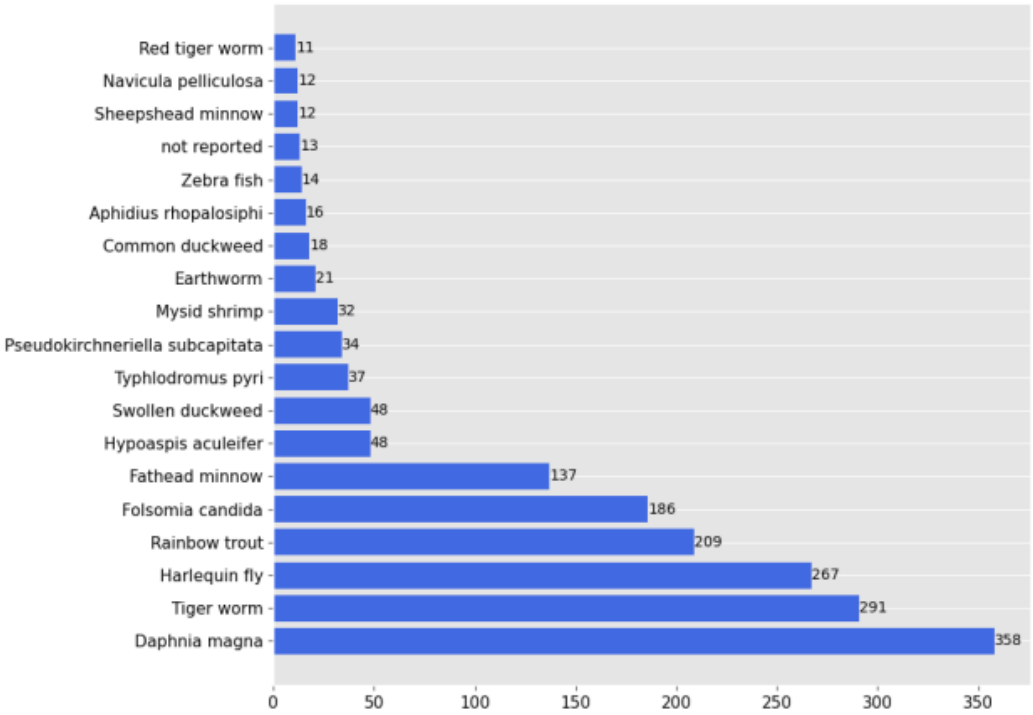

Figure 9. Classification of toxicity studies in OpenFoodTox for ecotox. Data entries in the database are shown for chronic/long term toxicity studies for reported species with sample size > 10.

The species with the most studies reported in the database are honey bee, daphnia magna, tiger worm and rainbow trout.

The latest version of OFT (released on 13 September 2023) was enriched with physicochemical and toxicokinetic properties and values on absorption, bioaccumulation, bioavailability, distribution and excretion. For the physicochemical

properties, data were collected from 605 EFSA outputs, covering 969 substances. Experimental data were available for 27 property types. For the toxicokinetic properties, data were collected from 577 EFSA outputs, covering 852 substances (**Table 2**). Thousands of values are present, providing a good level of detail of the behaviour of the substances in the body.

Table 3 shows the number of data and the properties relative to the physicochemical properties. Solubility is the most represented property, with thousands of values relative to both solvent and water. The appearance, physical status and colour data are abundant too, with thousands of data. UV/vis absorption and partition coefficients have more than 2000 values.

Table 2. Number of data for different toxicokinetic parameters.

parameters	**number of data**
absorption	580
bioaccumulation	473
bioavailability	64
distribution	1948
excretion	2222
toxicokinetics	2345
Total	7632

Table 3. Number of data for different physicochemical properties.

properties	**number of data**	**properties**	**number of data**
appearance/physical state/colour	2250	oxidising properties, other	391
boiling point	691	partition coefficient	1006
bulk density	42	refractive index	243
density, other	44	relative density	305
dissociation constant	528	relative self-ignition temperature (solids)	1
explosiveness, other	447	solubility in organic solvents/fat solubility	3902
flammability, other	418	surface tension	483
flammable gases	1	tap density	4
flammable solids	6	temperature of decomposition (state purity)	463
flash point, other	48	uv/vis absorption	2166
Henry's law constant	561	vapour pressure	681
melting point/freezing point	663	viscosity	54
optical rotation	7	water solubility	1572
Total 16,977			

OFT was used together with other relevant databases (e.g., US-EPA terrestrial database, Fraunhofer RepDose) to collect data for the development of several *in silico* models during the OptiTOX project. These models were developed for a range of test species and endpoints of relevance to human health, animal health and ecological risk assessment [4–10]. Most of the models (for instance related to bee toxicity, NOAEL

values, earthworms, etc.) were also implemented in the open-source VEGA platform[6] and are freely available [11].

3. Future perspective: The new OFT 3.0 database

The first version of OFT was published in 2017; since then, the database has been continuously updated and improved with new data related to several fields and properties.

The Chemical Strategy for Sustainability (CSS) proposed by the EU commission in 2020 is an important initiative of the European Green Deal.

The CSS aims to:

- achieve a toxic-free environment and a digital transition of the chemical sector;
- promote innovation with an increase in the use of alternative methods in the Risk Assessment (RA).

A pivotal point of the CSS is to remove legislative obstacles to the re-use of data and to simplify the exchange of chemical data between EU and national authorities. Moreover, it aims to achieve interoperability by ensuring that the data are shared using appropriate formats and tools (e.g., using IUCLID to store and disseminate hazard data).

To follow the actions outlined by the CSS, the European Chemicals Agency (ECHA) and EFSA published a joint paper [12] that has the key objective of establishing a simpler "one substance, one assessment (OSOA)" process for assessing the risks and hazards of chemicals. The main goal of OSOA is to ensure that all agencies in EU will have access to the same data, so that it will become possible to perform the assessment of the same substance in a harmonised way. The OSOA should lead to a simpler, faster and more transparent process of risk assessment, as a result of striving to have all available data in the same structured format. On this regard, the European Commission established a working group to work on:

- The use of IUCLID and the IUCLID format to collect hazard data;
- The establishment of the Common data platform on chemicals, which will provide a single access point to all data and information on chemicals in EU for all authorities. This will ensure that all the authorities have access to the same chemical data and each other's regulatory actions, and that all the stakeholders can have easy access to the same and best tools available for assessment;
- The establishment of an EU repository of health and environmental-based limit values (HEBLV), as part of the Common data platform, which aims to collect threshold-based limit values (ADI, AOEL etc.), legislation related to the setting of the limit values and also limit values derived by regulators (e.g., maximum residue level - MRL).

The EU Commission stated that OFT could be used as a starting point for the creation of the HEBLV repository.

In line with the OSOA approach as part of the CSS, EFSA proposed a new project (OC/EFSA/IDATA/2022/02) with the aim of further improving data quality and interoperability of EFSA's OFT database with IUCLID 6 and the EU Common data platform on chemicals, and of enhancing its usability as a supporting tool for risk assessment activities.

Four objectives are outlined in the new project, which is called sOFT-ERA:

- Objective 1: Update of the OpenFoodTox (OFT) database
- Objective 2: Further development of OFT 3.0
- Objective 3: Development of *in silico* models and implementation within EFSA tools
- Objective 4: Establishment of a process workflow for the integration of hazard data into IUCLID/OFT 3.0 as a part of the EFSA outputs publication

3.1. Objective 1. Update of the OpenFoodTox database

As previously described, the EU Common data platform on chemicals is currently under development. The OFT 3.0 project started with an analysis of the data content of OFT and mapping to the IUCLID schema, to ensure the proper migration of OFT data according to the IUCLID format (i.e., OECD Harmonised Templates). Most of the data currently contained in OFT version 6 (13 September 2023) have been migrated to IUCLID using IUCLID Data Uploader[7].

The migration of the data to the new format required a quality check. The technical description of the adopted strategy and the work done will be provided as an appendix.

IUCLID (International Uniform ChemicaL Information Database) was co-developed by ECHA and by the OECD. IUCLID is a software application that can be used to prepare, store, maintain and exchange hazard properties of chemical substances or mixtures and the associated exposure data. Under the REACH regulation, IUCLID was selected as the tool to be used by regulatory authorities and the chemical industry for data collection and submission [13].

By using OECD Harmonised Templates, all the data on chemicals in IUCLID can be captured and reported in a structured and harmonised way using the same format[8].

3.2. Objective 2: Further development of OFT 3.0

The sOFT-ERA project intends to further develop OFT 2.0 so that the updated OFT 3.0 can become one of the foundations for the EU's OSOA initiative. To achieve this, the IUCLID format represents a key pillar. However, this is not sufficient. The different EC regulations include many different endpoints. Thus, also a complete check of the ontologies related to the different endpoints has to be done.

The current version of OFT already contains data related to physicochemical and toxicokinetic properties, as previously described. During this four-year project the collection of data will continue using the appropriate OHTs as reference. For example, physicochemical properties and bioaccumulation data will be represented using the appropriate OHT controlled vocabularies (OHT 1 - 23 and 32 - 33). The new approach will be to collect not only data related to properties already collected in the last version of OFT, but also new types of properties.

These new additional properties are related to exposure and in vitro and in vivo toxicokinetic data relevant for human, animal health and environmental risk assessment. To cover this kind of data, further protocols are necessary, and additional OHTs: OHT 301 to 306 for in vitro and in vivo toxicokinetics; OHT 201 for

intermediate effects. The latter template is useful for reporting in vitro and *in silico* data.

Furthermore, data from other databases (e.g., ECHA; FAO/WHO data; EPA CompTox Dashboard) will be used to fill the existing data gaps.

Another kind of data which is being addressed within the new project is related to structured in vivo critical toxicological data, such as reference points/points of departure and reference values relevant for human health risk assessment from FAO/WHO (i.e., JECFA and JMPR outputs). In this particular case, the OHTs of reference are: OHT 58 to 86 for health effects, and OHT 201 for intermediate effects. The integration of content from the Joint FAO/WHO Expert Committee on Food Additives and the Joint FAO/WHO Meeting on Pesticide Residues (JECFA/JMPR) will be extremely beneficial.

sOFT-ERA will also collect and curate in vivo and in vitro toxicological dose response data of relevance to human health, animal health and environmental risk assessment. Regarding the OHTs, which shall be used, they are: OHT 58 to 84, and 86 for health effects, OHT 41 to 57 for effects on biotic systems, and OHT 201 for intermediate effects. This activity will also be useful to retrieve data associated with mechanistic studies.

As additional types of values, sOFT-ERA will collect and integrate structured in vivo critical and non-critical toxicological data relevant for human health risk assessment. In this case, the reference OHTs are: OHT 58 to 84, and 86 for health effects, and OHT 201 for intermediate effects. Whilst the data content of OFT 2.0 is extremely informative and reliable for understanding chemical safety with its critical effects and studies applied to the HEBLVs, more in-depth information for other non-critical effects would also be desirable. This information could further the understanding of critical effects, and anticipate their appearance. These actions could contribute towards having a much more sophisticated and accurate risk assessment in the future.

Within this more complex framework, it is important to note that the target of the effort is not only related to the prediction of a toxic effect, for instance, but also to the use of this information for other purposes, for instance to aggregate substances for the establishment of cumulative assessment groups (CAGs). In this context, data within OFT could be exploited to group substances using common features.

3.3. Objective 3: Development of *in silico* models and implementation within EFSA tools

An important part of the project is the development of new *in silico* models using the data collected in OFT. The modelling tools will address not only prediction of specific endpoints, but also models for grouping, to be used for CAG. Furthermore, models to be used for read-across will also be developed.

Thus, the development of *in silico* tools is expected to follow three different scenarios:

a. Predicting property values, in case of missing values. Ideally this can complete the data matrix, filling cells with missing values. However, some considerations are necessary: the source of the value has to be clear (i.e., if the value results from

an experiment or from an *in silico* model); if the predicted values have higher uncertainty compared with experimental value, it is important to distinguish the level of uncertainty depending on the substance and the model; the models which will be developed will allow to determine the specific level of uncertainty, and the approach is what has been described using the applicability domain index. It was demonstrated that this approach is efficient to differentiate if the prediction is more reliable [14]. Based on this kind of consideration, EFSA has already introduced the criterion to use the information on the reliability of the individual prediction. This approach is consistent with the EFSA Guideline about Weight-of-evidence [1], which specifies the procedure to integrate data from multiple sources, including read-across and predictive *in silico* models.

b. Use of *in silico* models to group substances. This is useful in the perspective of the identification of CAGs. One example of application is to cluster substances for co-exposure, where, according to the EFSA Guidance on co-exposure [15], substances within the same group can be addressed jointly within the Dose Addition approach related to mixture. If the user does not have the information on the necessary toxicological pathway (which may be provided in different ways, such as mechanism of action, mode of action, adverse outcome pathway), the use of *in silico* models can cover this gap of knowledge. It is clear that in this case, the model is not used directly for the evaluation of the individual adverse effect; the prediction serves to identify families of substances that share the same toxicological profile, and thus the doses of these substances can be added.

c. Read-across is a third independent way to process chemical substances. Traditionally, read-across has been done manually by experts. Within the previous project related to OFT, an innovative read-across tool called VERA (Virtual Extensive Read-Across) was developed, which identifies the relevant similar substances, using structural similarity (as in the VEGA software), molecular groups and structural alerts [16,17]. These components are used to define families of similar substances, and then the software compares these families. The VERA software is freely available on the VEGAHUB website[9]. Within the new project, further metrics for similarities will be introduced, and the application will also be extended to several endpoints. The use of read-across is useful both to cover data gaps, but also to analyse more in depth the commonalities between different substances, if the software allows to perform the comparison using different metrics. In other words, this tool will not only provide a list of similar substances, but also the explicit elements useful to hypothesize about the differences between two substances and the role of specific components, which may increase or decrease the effect. In this way, the evaluation will become more transparent and useful, providing information on specific molecular features and their contribution to the final effect.

3.4. Objective 4: Establishment of a process workflow for the integration of hazard data into IUCLID/OFT 3.0 as a part of the EFSA outputs publication

One of the issues that will be addressed in this project is the delay between data

collection and data publication in OFT. To date, OFT is updated each year with new data from EFSA documents.

Objective 4 aims at creating a workflow that will allow the integration of the updates of OFT as part of the EFSA output publication into IUCLID/OFT 3.0. In this way, it will be possible, not to eliminate but to reduce the time gap between the publication of the EFSA outputs and the collection of these in OFT.

4. Conclusion

The aims of the CSS and the OSOA approach are to have a single access point to data and information on chemicals in EU for all authorities (i.e., the Common data platform on chemicals); to achieve a simpler exchange of chemical data between EU and national authorities; to ensure interoperability (i.e., it is necessary to have all data in the same format, and this will be made possible through the use of the IUCLID format). Within this context, OFT was selected as the starting point for the creation of the HEBLV repository as part of the Common data platform on chemicals.

Since his creation, OFT has been updated with more than 5000 unique substances using more than 2400 EFSA documents (opinions, statements, and conclusions). It is open-source and available for download on the OFT MicroStrategy dashboard as a full database or as specific dataset/datasheet. To follow the roadmap of the CSS, the last version of OFT (13 September 2023) will be mapped to the IUCLID format and migrated to IUCLID. Moreover, OFT will be further updated by collecting new data. The data collected in OFT have been used to develop *in silico* models and will continue to provide models for new endpoints and species.

Author contributions: Conceptualization, NI, EB, AR, LS and CY; writing—original draft preparation, NI, EB, AR, LS and CY; writing—review and editing, NI, EB, AR, LS and CY. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the financial contribution within the project sOFT-ERA, OC/EFSA/IDATA/2022/02.

Conflict of interest: The authors declare no conflict of interest.

Notes

1 http://ambit.sourceforge.net
2 https://comptox.epa.gov/dashboard/
3 https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test
4 10.510.5281/zenodo.8120114281/zenodo.8120114
5 https://www.efsa.europa.eu/en/microstrategy/openfoodtox
6 https://www.vegahub.eu/
7 https://iuclid6.echa.europa.eu/it/data-uploader
8 https://echa.europa.eu/support/registration/creating-your-registration-dossier/what-is-iuclid-
9 www.vegahub.eu

References

1. EFSA Scientific Committee, Hardy A, Benford D, et al. Guidance on the use of the weight of evidence approach in scientific assessments. EFS2. 2017; 15(8). doi: 10.2903/j.efsa.2017.4971
2. Dorne JLCM, Richardson J, Livaniou A, et al. EFSA's OpenFoodTox: An open source toxicological database on chemicals in food and feed and its future developments. Environment International. 2021; 146: 106293. doi: 10.1016/j.envint.2020.106293
3. EFSA. EFSA Strategy 2020 Trusted science for safe food. EFSA; 2016.
4. Benfenati E, Como F, Manzo M, et al. Developing innovative in silico models with EFSA's OpenFoodTox database. EFS3. 2017; 14(7). doi: 10.2903/sp.efsa.2017.EN-1206
5. Como F, Carnesecchi E, Volani S, et al. Predicting acute contact toxicity of pesticides in honeybees (Apis mellifera) through a k-nearest neighbor model. Chemosphere. 2017; 166: 438–444. doi: 10.1016/j.chemosphere.2016.09.092
6. Carnesecchi E, Toma C, Roncaglioni A, et al. Integrating QSAR models predicting acute contact toxicity and mode of action profiling in honey bees (A. mellifera): Data curation using open source databases, performance testing and validation. Science of The Total Environment. 2020; 735: 139243. doi: 10.1016/j.scitotenv.2020.139243
7. Gadaleta D, Marzo M, Toropov A, et al. Integrated In Silico Models for the Prediction of No-Observed-(Adverse)-Effect Levels and Lowest-Observed-(Adverse)-Effect Levels in Rats for Sub-chronic Repeated-Dose Toxicity. Chemical Research in Toxicology. 2020; 34(2): 247–257. doi: 10.1021/acs.chemrestox.0c00176
8. Ghosh S, Ojha PK, Carnesecchi E, et al. Exploring QSAR modeling of toxicity of chemicals on earthworm. Ecotoxicology and Environmental Safety. 2020; 190: 110067. doi: 10.1016/j.ecoenv.2019.110067
9. Lavado GJ, Baderna D, Carnesecchi E, et al. QSAR models for soil ecotoxicity: Development and validation of models to predict reproductive toxicity of organic chemicals in the collembola Folsomia candida. Journal of Hazardous Materials. 2022; 423: 127236. doi: 10.1016/j.jhazmat.2021.127236
10. Toropov AA, Di Nicola MR, Toropova AP, et al. A regression-based QSAR-model to predict acute toxicity of aromatic chemicals in tadpoles of the Japanese brown frog (Rana japonica): Calibration, validation, and future developments to support risk assessment of chemicals in amphibians. Science of The Total Environment. 2022; 830: 154795. doi: 10.1016/j.scitotenv.2022.154795
11. Roncaglioni A, Lombardo A, Benfenati E. The VEGAHUB Platform: The Philosophy and the Tools. Alternatives to Laboratory Animals. 2022; 50(2): 121–135. doi: 10.1177/02611929221090530
12. EFSA. In support of the EU chemicals strategy for sustainability: One substance – one assessment. EFSA; 2020.
13. OECDilibrary. OECD Series on Testing and Assessment. Customisation Opportunities of IUCLID for the Management of Chemical Data – 3rd edition. OECDilibrary; 2023.
14. Danieli A, Colombo E, Raitano G, et al. The VEGA Tool to Check the Applicability Domain Gives Greater Confidence in the Prediction of In Silico Models. International Journal of Molecular Sciences. 2023; 24(12): 9894. doi: 10.3390/ijms24129894
15. European Food Safety Authority (EFSA), Charistou A, Coja T, et al. Guidance on the assessment of exposure of operators, workers, residents and bystanders in risk assessment of plant protection products. EFS2. 2022; 20(1). doi: 10.2903/j.efsa.2022.7032
16. Benfenati E, Roncaglioni A, Iovine N, et al. Maintenance, update and further development of EFSA's Chemical Hazards: OpenFoodTox 2.0. EFS3. 2024; 21(1). doi: 10.2903/sp.efsa.2024.EN-8590
17. Viganò EL, Colombo E, Raitano G, et al. Virtual Extensive Read-Across: A New Open-Access Software for Chemical Read-Across and Its Application to the Carcinogenicity Assessment of Botanicals. Molecules. 2022; 27(19): 6605. doi: 10.3390/molecules27196605